

## 附件 1：单变量模型

背景例子：一个狮鹫（西方世界传说中的一种神兽）种群，我们测量了该种群每个个体的出生体重，也知道了该种群的遗传家系信息。这两个数据表格详见附件（“gryphon.txt”；“gryphonped.txt”）。根据其数据，第一，可以计算该性状（出生体重）的遗传力；第二，因为也知道每个个体的母亲，我们在该模型中（混合模型）加入另一个随机因变量母亲的 ID，来计算该性状的母系遗传；第三，整合该种群的性别信息，将其作为一个固定的因变量，主要有两个作用，可以得出该性状在两性间是否有差异，同时在控制性别对出生体重的影响下，能更加精确地计算以上所述遗传参数。

### 1 程序包安装

MCMCglmm 为 R 语言的一个程序包，因此第一步则是在 R 语言某一个镜像下下载并安装程序包 MCMCglmm，因为涉及到遗传家系，R 语言中有一个专用程序包“pedantics”（Morrissey 2015）处理家系，因此也一并安装该程序包。

### 2 数据读入及随机、固定变量设置

将本文所附件（“gryphon.txt”；“gryphonped.txt”）的 2 个文本文档，一个为狮鹫出生体重测量文档（gryphon.txt），另一个为该种群家系文档（gryphonped.txt）置于电脑上 R 语言的工作路径下。

R 代码（绿色）：

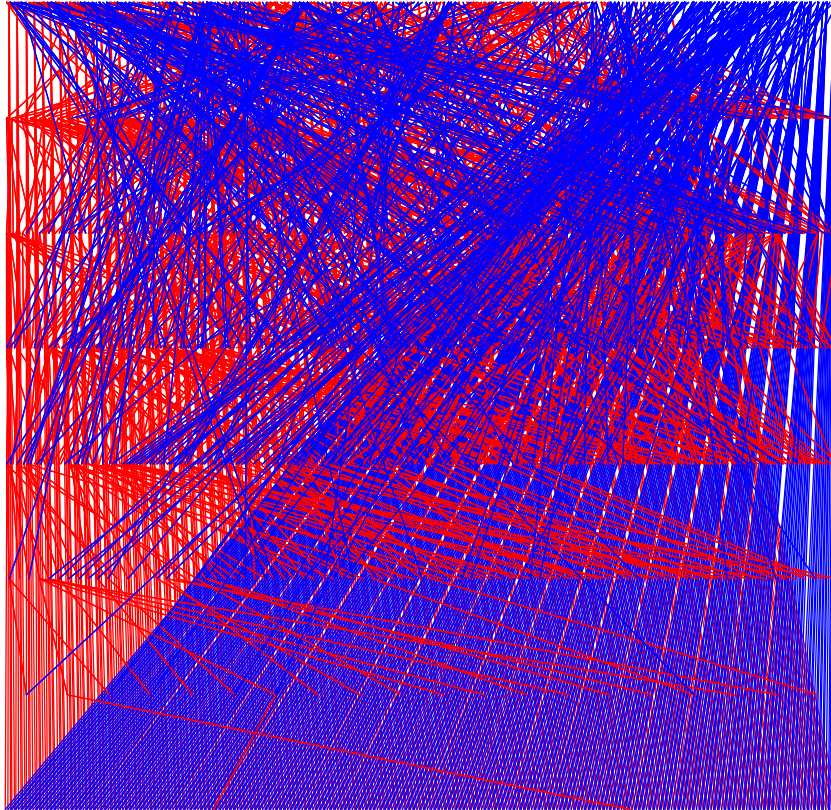
```
getwd () # 确认你电脑 R 语言的默认工作路径，将上述两个 txt 文件置于该工作路径下。  
library(pedantics) # 载入程序包“pedantics”  
library(MCMCglmm) # 载入程序包“MCMCglmm”  
gryphonPED<-read.table("C:\\data\\gryphonped.txt",header=T # 读入种群家系数据  
head(gryphonPED) # 显示该家系数据前 6 行，如下所示
```

| ID     | FATHER | MOTHER |
|--------|--------|--------|
| 1 1306 | NA     | NA     |
| 2 1304 | NA     | NA     |
| 3 1298 | NA     | NA     |
| 4 1293 | NA     | NA     |
| 5 1290 | NA     | NA     |
| 6 1288 | NA     | NA     |

该数据显示 3 列，第一列为个体 ID，第二列为母亲（DAM）ID，第三列为父亲（SIRE）ID。NA 表示缺失值（一般第一代个体不知道其父母亲）。

```
drawPedigree(gryphonP
```

#根据读入的数据画出整个种群的遗传家系图, 红色表示母系, 蓝色表示父亲。



```
gryphon<-read.table("C:\\data\\gryphon.txt",header=T) #读入性状表型值数据
```

```
head (gryphon ) #显示该数据前 6 行，如下所示
```

|   | ANIMAL | MOTHER | BYEAR | SEX | BWT   | TARSUS |
|---|--------|--------|-------|-----|-------|--------|
| 1 | 1029   | 1145   | 968   | 1   | 10.77 | 24.77  |
| 2 | 1299   | 811    | 968   | 1   | 9.30  | 22.46  |
| 3 | 643    | 642    | 970   | 2   | 3.98  | 12.89  |
| 4 | 1183   | 1186   | 970   | 1   | 5.39  | 20.47  |
| 5 | 1238   | 1237   | 970   | 2   | 12.12 | NA     |
| 6 | 891    | 895    | 970   | 1   | NA    | NA     |

该数据表格第一列为个体 ID “ANIMAL”，第二列为母亲 ID，第四列为性别，第五列为出生体重，是我们所需要的数据。因为 R 语言对于字母大小写很敏感，而 MCMCglmm 模型代码中一般都将随机变量特定为“animal”，因此有：

```

gryphon$animal <-as.factor(gryphon$ANIMAL
# 将第一列变量名称改成小写同时因子化
gryphon$MOTHER <-as.factor(gryphon$MOTHER)
# 将母亲 ID“MOTHER”因子化作为另一个随#机变量
gryphon$SEX <-as.factor(gryphon$SEX)
# 将性别“因子化”，固定变量

```

### 3 模型运行代码和结果

```

prior1 <- list(G = list(G1 = list(V = 1, n = 0.002), G2 = list(V = 1, #先验设置
n = 0.002)), R = list(V = 1, n = 0.002))

```

```

Model1 <- MCMCglmm(BWT ~ SEX, random = ~animal + MOTHER,
family = c("gaussian"), #假设因变量出生体重“BWT”为正态分布
pedigree = gryphonPED, data = gryphon , prior = prior1)

```

#具体模型代码：出生体重“BWT”为因变量，特定随机变量为“animal”，动物模型就是以这个特定的随机变量作为两个数据（家系数据和表型值数据）的桥梁来计算该性状的遗传力。另一个随机变量为母亲 ID（MOTHER），在该模型中我们加入这个随机变量用来计算该性状的母系遗传效应。注意，我们还加入了性别“SEX”，视为固定变量，用来比较雄性和雌性出生体重的差异。“pedigree= gryphonPED”表示这个模型需要的家系数据来自读入的文档“gryphonPED”，与之类似，“data= gryphon”表示这个模型所需的表型值数据读入的文档“gryphon”，最后“prior=prior1”表示该模型的先验采用先前设置的 prior1。

```
summary(Model1)
```

#显示模型运行结果 重要结果如下（符号“#”后为中文为解释）：

```
Iterations = 3001:12991 Thinning interval = 10 Sample size = 1000 DIC: 3566.029
```

```
G-structure: ~animal
```

#表示  $V_G$  部分，这里具体来说是包括加性遗传变异和母系效应部分

|        | post.mean | l-95% CI | u-95% CI | eff.samp |
|--------|-----------|----------|----------|----------|
| animal | 2.72      | 1.654    | 3.898    | 89.11    |

# 加性遗传变异绝对值±95% CrI

```
~MOTHER
```

# 母性遗传效应绝对值±95% CrI

|  | post.mean | l-95% CI | u-95% CI | eff.samp |
|--|-----------|----------|----------|----------|
|--|-----------|----------|----------|----------|

```
MOTHER 1.064 0.5952 1.579 330
```

```
R-structure: ~units
```

```
#表示  $V_R$  部分，这里具体来说是除加性遗传和母系效应部分外的其他残差
```

|       | post.mean | l-95% CI | u-95% CI | eff.samp |
|-------|-----------|----------|----------|----------|
| units | 2.296     | 1.383    | 3.119    | 97.93    |

```
Location effects: BWT ~ SEX
```

```
#显示固定变量性别“SEX”的结果
```

|             | post.mean | l-95% CI | u-95% CI | eff.samp | pMCMC      |
|-------------|-----------|----------|----------|----------|------------|
| (Intercept) | 6.079     | 5.738    | 6.402    | 1000     | <0.001 *** |
| SEX2        | 2.214     | 1.904    | 2.506    | 1099     | <0.001 *** |

```
# 雄性出生体重显著大于雌性
```

至此，我们已经得出所有的变量值，对于遗传力，只需计算其相对比重即可得出：

```
posterior.heritability <- Model1$Vcov[, "animal"]/(Model1$Vcov[, "animal"] + Model1$Vcov[,  
"MOTHER"] + Model1$Vcov[, "units"])
```

```
#计算加性遗传变异比重：遗传力
```

```
posterior.mode(posterior.heritability)
```

```
# 遗传力平均值（贝叶斯 1000 个后验遗传力的平均值）
```

此次（每次运行，尤其贝叶斯模拟和取样，可能会有细微的差别）得出该性状（出生体重）的遗传力为：0.4661521

```
posterior.maternal <- Model1$Vcov[, "MOTHER"]/(Model1$Vcov[, "animal"] + Model1$Vcov[,  
"MOTHER"] + Model1$Vcov[, "units"]) #计算该性状母系效应
```

```
posterior.mode(posterior.maternal) # 母性效应平均值
```

此次得出该性状（出生体重）的母性效应为：0.1723112

## 附件 2: 多 (双) 变量模型

背景例子: 一个学生想知道生长速率和免疫功能的关系, 这两者是否存在一个 trade-off? 他/她以野外一个田鼠种群作为研究对象。做出这样的预测: 这两个性状都是可遗传而且受到自然选择的正驱动, 种群在快速的生长率 (growth rate: GR) 和更好的免疫功能 (Immune function: IF) 这两者间有一个权衡, 因此会有一个负遗传相关关系。为了验证这个假说, 他调查了该种群的遗传家系 (可以利用亲本分析加以辅助), 同时测量了每个个体的这两个性状表型值, 利用 MCMCglmm 来运行动物模型。

### 1 程序包安装

见附件 1

### 2 数据读入

将本文所附件的 2 个文本文档, 一个为田鼠表型测量文档 (volesDATA.txt), 一个为该种群家系文档 (volesPED.txt) 置于你电脑上 R 语言的工作路径下。

R 代码 (绿色):

```
getwd ()
```

```
# 确认你电脑 R 语言的默认工作路径, 将上述两个 txt 文件置于该工作路径下。
```

```
library(pedantics)
```

```
# 载入程序包“pedantics”
```

```
library(MCMCglmm)
```

```
# 载入程序包“MCMCglmm”
```

```
volesPED<-read.table("C:\\data\\volesPED.txt",header=T)
```

```
# 读入种群家系数据
```

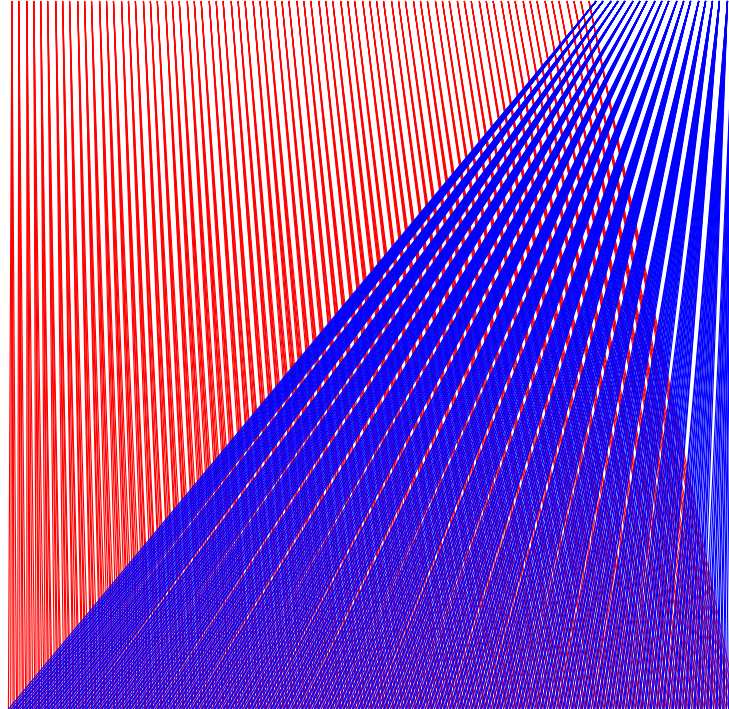
```
head(volesPED)
```

```
# 显示该家系数据前 6 行, 如下所示
```

```
  ID DAM SIRE
1 401 NA  NA
2 402 NA  NA
3 403 NA  NA
4 404 NA  NA
5 405 NA  NA
6 406 NA  NA
```

该数据显示 3 列，第一列为个体 ID，第二列为母亲（DAM）ID，第三列为父亲（SIRE）ID。  
NA 表示缺失值（一般第一代个体不知道其父母亲）。

`drawPedigree(volesPED)` #根据读入的数据画出整个种群的遗传家系图，如下：



由于是模拟数据，因此该家系图非常简单：红色表示母系，蓝色表示父亲。

```
voles<-read.table("C:\\data\\volesDATA.txt", header=T)
```

```
#读入性状表型值数据
```

```
voles$ID<-as.factor(voles$ID)
```

```
#将个体ID“因子化”，随机变量
```

```
voles$SEX<-as.factor(voles$SEX)
```

```
#将性别“因子化”，固定变量
```

```
head(voles)
```

```
#显示该数据前 6 行，如下所示
```

|     | ID  | RA       | GR        | IF       | SEX | w         |
|-----|-----|----------|-----------|----------|-----|-----------|
| 401 | 401 | 2.933001 | 7.465881  | 5.652952 | 1   | 0.9575802 |
| 402 | 402 | 5.049779 | 10.045831 | 5.470318 | 1   | 1.0780310 |
| 403 | 403 | 3.897006 | 9.017215  | 5.218483 | 1   | 1.0149944 |
| 404 | 404 | 1.665640 | 7.850139  | 3.599462 | 1   | 0.8942304 |
| 405 | 405 | 5.333196 | 9.274917  | 7.216384 | 1   | 1.0953102 |

```
406 406 4.731553 9.274813 6.160116 1 1.0609742
```

该数据第一列为个体 ID，第二列为另一个变量，第三列为生长速率（GR），第四列为免疫功能（IF）。

```
voles$animal<-as.factor(voles$ID)
```

```
#将数据“voles”的第一列变量名称从“ID”改成
```

```
#“animal”：MCMCglmmcheng 程序包自动识别“animal”
```

### 3 模型运行代码和结果

```
prior 1<- list(G = list(G1 = list(V = diag(2), n = 1.002)), #设置先验参数
```

```
      R = list(V = diag(2), n = 1.002))
```

```
model1 <- MCMCglmm(cbind(GR,IF) ~ trait - 1+trait:SEX,
```

```
      random = ~us(trait):animal, rcov = ~us(trait):units,
```

```
      family = c("gaussian", "gaussian"),
```

```
      pedigree = volesPED, data = voles, prior = prior1)
```

```
# 模型代码，因变量为“GR”和“IF”# 将SEX视为固定变量
```

```
# 随机变量为animal
```

```
#假设 GR和IF皆为正态分布
```

```
#“voles”，先验为之前设置的“prior1”
```

```
#家系为“volesPED”，数据为
```

```
summary(model1) #显示模型运行结果 重要结果如下（中文为解释）：
```

```
Iterations = 3001:12991      Thinning interval = 10, Sample size = 1000      DIC: 2439.624
```

```
G-structure: ~us(trait):animal
```

```
#表示  $V_G$  部分，这里具体来说是包括两个性状的加性遗传变异部分
```

|                        | post.mean | 1-95% CI | u-95% CI | eff.samp |
|------------------------|-----------|----------|----------|----------|
| traitGR:traitGR.animal | 0.4435    | 0.2817   | 0.63593  | 225.7    |

```
#生长速率加性遗传绝对值
```

|                        |         |         |         |       |
|------------------------|---------|---------|---------|-------|
| traitIF:traitGR.animal | -0.1145 | -0.2275 | 0.02839 | 209.7 |
|------------------------|---------|---------|---------|-------|

```
#加性协方差绝对值
```

|                        |         |         |         |       |
|------------------------|---------|---------|---------|-------|
| traitGR:traitIF.animal | -0.1145 | -0.2275 | 0.02839 | 209.7 |
|------------------------|---------|---------|---------|-------|

```
#同上
```

|                        |        |        |         |       |
|------------------------|--------|--------|---------|-------|
| traitIF:traitIF.animal | 0.3995 | 0.2088 | 0.57763 | 218.8 |
|------------------------|--------|--------|---------|-------|

#免疫功能假性遗传绝对值

R-structure: ~us(trait):units

#表示  $V_R$  部分，这里具体来说是除加性遗传部分外的其他残差

|                       | post.mean | l-95% CI | u-95% CI | eff.samp |
|-----------------------|-----------|----------|----------|----------|
| traitGR:traitGR.units | 0.7819    | 0.6317   | 0.9316   | 379.9    |

#生长速率残差绝对值

|                       |        |        |        |       |
|-----------------------|--------|--------|--------|-------|
| traitIF:traitGR.units | 0.6498 | 0.5291 | 0.7777 | 467.7 |
|-----------------------|--------|--------|--------|-------|

#残差协方差绝对值

|                       |        |        |        |       |
|-----------------------|--------|--------|--------|-------|
| traitGR:traitIF.units | 0.6498 | 0.5291 | 0.7777 | 467.7 |
|-----------------------|--------|--------|--------|-------|

#同上

|                       |        |        |        |       |
|-----------------------|--------|--------|--------|-------|
| traitIF:traitIF.units | 0.8260 | 0.6513 | 0.9722 | 419.5 |
|-----------------------|--------|--------|--------|-------|

#免疫功能残差绝对值

Location effects: cbind(GR, IF) ~ trait - 1 + trait:SEX

#表示固定变量性别的结果

|              | post.mean | l-95% CI | u-95% CI | eff.samp | pMCMC      |
|--------------|-----------|----------|----------|----------|------------|
| traitGR      | 8.0786    | 7.9045   | 8.2420   | 1000.0   | <0.001 *** |
| traitIF      | 4.8813    | 4.7256   | 5.0618   | 612.7    | <0.001 *** |
| traitGR:SEX2 | 0.9532    | 0.7679   | 1.1258   | 1000.0   | <0.001 *** |
| traitIF:SEX2 | -0.5462   | -0.7070  | -0.3435  | 1000.0   | <0.001 *** |

从上述中我们可以得出，雄性个体生长速率显著高于雌性，但是免疫功能显著低于雌性。

现在我们根据上面的各变量相对差来计算遗传相关：

```
G<- matrix(data=posterior.mode(model1$VCV)[1:4], nrow=2, ncol=2)
```

```
G
```

得出一下矩阵，我们可以看到 GR 和 IF 是负遗传相关的

```
      [,1] [,2]
[1,] 0.4467877 -0.1331386
[2,] -0.1331386 0.3700549
```

以上矩阵表示遗传相关绝对值，让我们标准化（把相关系数标准尺度：从负 1 到正 1）：

```
Gcor<-cov2cor(G)
```

```
Gcor
```



得出一下矩阵，更易理解，GR 和 GR 自己的相关系数为 1，和 IF 具有很强的负遗传相关，为-0.33.

```
      [,1] [,2]
[1,] 1.0000000 -0.3274313
[2,] -0.3274313 1.0000000
```