Phantom Groups

LRS

CGIL

July-Aug 2012

Incomplete Pedigrees

- Parent IDs are often missing.
    - Animals are transfered to new owners.
    - Animal IDs are changed (horses).
    - On/Off Recording programs.
    - Multiple sires, unattended births.
- Not all unknown parents belong to the same population of animals.

How To Group

                              Animal
                  Male   Female
           SM   DM   SF   DF

Then each of these are nested within year of birth of the animal.

Example

Animal	Sire	Dam	Gender	BYear
A			M	2008
B			F	2009
C			F	2009
D	A		M	2009
E		B	M	2009

Example

Animal	Sire	Dam	Gender	BYear
A	SM8	DM8	M	2008
B	SF9	DF9	F	2009
C	SF9	DF9	F	2009
D	A	DM9	M	2009
E	SM9	B	M	2009

Every animal has some parent identification after grouping.

Group Model

$$\mathbf{y} = \mathbf{Xb} + \mathbf{ZQg} + \mathbf{Za} + \mathbf{e},$$

where

      $\mathbf{a}$ is the vector of animal additive genetic effects,

      $\mathbf{Z}$ is the matrix that relates animals to their observations,

      $\mathbf{g}$ is the vector of genetic group effects, and

      $\mathbf{Q}$ is the matrix that relates animals to their genetic groups, and

  $\mathbf{y}, \mathbf{Xb}, \mathbf{e}$ are as described in earlier notes.

The *Estimated Breeding Value*, EBV, of an animal is equal to the sum of the group and animal solutions from MME, i.e.

$$\text{Vector of EBVs} = \mathbf{Q}\hat{\mathbf{g}} + \hat{\mathbf{a}}$$

Q Matrix

Animal	Row of **Q**					
	SM8	DM8	SF9	DF9	SM9	DM9
A	.5	.5	0	0	0	0
B	0	0	.5	.5	0	0
C	0	0	.5	.5	0	0
D	.25	.25	0	0	0	.5
E	0	0	.25	.25	.5	0

Q Matrix

Animal			Row of $\mathbf{Q}$			
	SM8	DM8	SF9	DF9	SM9	DM9
A	.5	.5	0	0	0	0
B	0	0	.5	.5	0	0
C	0	0	.5	.5	0	0
D	.25	.25	0	0	0	.5
E	0	0	.25	.25	.5	0
F(E,D)	.125	.125	.125	.125	.25	.25

Every animal has some function of group effects in $\mathbf{Q}$.

MME

$$\begin{pmatrix} X'X & X'ZQ & X'Z \\ Q'Z'X & Q'Z'ZQ & Q'Z'Z \\ Z'X & Z'ZQ & Z'Z + A^{-1}\alpha \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{g} \\ \hat{a} \end{pmatrix} = \begin{pmatrix} X'y \\ Q'Z'y \\ Z'y \end{pmatrix}$$

Q' times the third row subtracted from the second row gives

$$Q'A^{-1}\hat{a}\alpha = 0$$

Setting up Q painful and messy.

Quaas-Pollak Transformation, 1981

$$\begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \\ \hat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{Q} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \\ \hat{\mathbf{a}} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{Q} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \\ \mathbf{Q}\hat{\mathbf{g}} + \hat{\mathbf{a}} \end{pmatrix}$$

QP Transformation

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & -\mathbf{Q}' \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{X'X} & \mathbf{X'ZQ} & \mathbf{X'Z} \\ \mathbf{Q'Z'X} & \mathbf{Q'Z'ZQ} & \mathbf{Q'Z'Z} \\ \mathbf{Z'X} & \mathbf{Z'ZQ} & \mathbf{Z'Z} + \mathbf{A}^{-1}\alpha \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{Q} & \mathbf{I} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{X'X} & \mathbf{0} & \mathbf{X'Z} \\ \mathbf{0} & \mathbf{Q'A}^{-1}\mathbf{Q}\alpha & -\mathbf{Q'A}^{-1}\alpha \\ \mathbf{Z'X} & -\mathbf{A}^{-1}\mathbf{Q}\alpha & \mathbf{Z'Z} + \mathbf{A}^{-1}\alpha \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \\ \mathbf{Q}\hat{\mathbf{g}} + \hat{\mathbf{a}} \end{pmatrix}$$

and

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & -\mathbf{Q}' \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{X'y} \\ \mathbf{Q'Z'y} \\ \mathbf{Z'y} \end{pmatrix} = \begin{pmatrix} \mathbf{X'y} \\ \mathbf{0} \\ \mathbf{Z'y} \end{pmatrix}$$

Quaas, 1988

- Dick Quaas noticed that $\mathbf{Q'A^{-1}Q}$ and $-\mathbf{Q'A^{-1}}$ had properties that followed Henderson's rules for forming $\mathbf{A^{-1}}$.
- Calculate b_{ii} values for all animals in pedigree as already discussed.
- Add phantom groups to the pedigree. Let their $b_{ii} = 1$.
- Create $\mathbf{A^{-1}}$ for all real animals, but allow for phantom groups to be part of $\mathbf{A^{-1}}$.

Example

	A	D	SM8	DM8	SM9
A	1		-.5	-.5	
D					
SM8	-.5		.25	.25	
DM8	-.5		.25	.25	
SM9					

Example

	A	D	SM8	DM8	SM9
A	$1\frac{1}{3}$	$-\frac{2}{3}$	-.5	-.5	$\frac{1}{3}$
D	$-\frac{2}{3}$	$\frac{4}{3}$			$-\frac{2}{3}$
SM8	-.5		.25	.25	
DM8	-.5		.25	.25	
SM9	$\frac{1}{3}$	$-\frac{2}{3}$			$\frac{1}{3}$

Computing Issues

- Group effects are fixed effects, rank could be less than number of groups (confounding). Sometimes very large solutions for group effects.
- Overlapping groups to break confounding.
- Pretend groups are animals with $b_{ii} = 1$, then add 1 to group diagonals of $\mathbf{A}^{-1}$. Group solutions then add to zero, dependencies removed.

Multiple Sires

In some species females are housed with several males, so that the sire at mating is unknown without blood testing sires and progeny. How can **A** and $\mathbf{A}^{-1}$ be modified?

Suppose female X was kept with sires A and B, each equally likely to have sired the progeny, Z.

	A	B	X	Z
A	1	0	0	$\frac{1}{4}$
B	0	1	0	$\frac{1}{4}$
X	0	0	1	$\frac{1}{2}$
Z	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	1

Simulation of Data

Simulation is one way of improving one's understanding of a model, a data set, a method, or a system.

$$y_{ijk} = Y_i + C_j + a_k + e_{ijk}$$

where

y_{ijk} is the observed trait, one record per animal,

Y_i is a fixed year effect,

C_j is a random contemporary group effect,

a_k is a random animal additive genetic effect,

e_{ijk} is a random residual effect.

Details

- Years $= 10$
- Contemporary groups $= 50$ to 100 per year
- $\sigma_e^2 = 64 = 1.6 \; \sigma_a^2 = 4 \; \sigma_c^2$
- Number of animals, up to 20000
- Base animals $= 200$
- Number of breeding animals per year
- 5 to 15 animals per CG

Initialization

```
zlrs = file.choose()  # find lrscrips.R
source(zlrs)
zbill = file.choose() # find Bills.R
zrclib = file.choose()  # find rclib.dll
dyn.load(zrclib)
source(zbill)
NAM = 20000        # max number of animals
nyr = 10           # number of years
SDa = sqrt(64/1.6)
SDc = sqrt(64/4)
SDe = 8
mbase = 200
tbv = rep(0,NAM) # also  inbb, inbc, sid, did,
obs, year, cgrp, anwr, bi, fi, atbv, sel
```

Initiate

```
xpdint(NAM)
YRFX = c(50, 53, 52, 55, 54, 51, 57, 56, 58, 59)
for(i in 1:mbase){
  ks=0; kd=0
  inbc[i] = xpdadd(ks,kd)
  inbb[i]=1
  }
sel[1:mbase] = 50 + rnorm(mbase,0,SDa)
aid = c(1:NAM)
nba = 200; nanim = mbase; nrec=0
cgid = 0
```

Year Loop

```
for(iyr in 1:nyr){
  # how many CG in this year?
  ncg = sample(c(50:100),1)
  ka = order(-sel)  # order all animals
  B = aid[ka]
  B = B[1:nba]  # keep the best for breeding

  Loop for CG

}
```

CG Loop

```
for(icg in 1:ncg){
  # How many animals in CG?
  mams = sample(c(5:15),1)
  cgfx = rnorm(1,0,SDc)
  cgid = cgid + 1

  Loop for Animals

}
```

Animal Loop

```
for(k in 1:mams){
  # Pick parents, any two in B
  nanim = nanim + 1
  nrec = nrec + 1
  parents = sample(B,2,replace=FALSE)
  ksir = parents[1]; kdam=parents[2]
  inbc[nanim] = xpdadd(ksir,kdam)
  inbb[nanim] = xpdd(nanim)
  x = sqrt(inbb[nanim])*SDa
  tbv[nanim] = 0.5*(tbv[ksir]+tbv[kdam])+rnorm(1,0,x)

  # Make the record
```

Record Making

```
# Make the record
  obs[nrec]=YRFX[iyr]+cgfx+tbv[nanim]+rnorm(1,0,SDe)
  year[nrec] = iyr;      cgrp[nrec] = cgid
  anwr[nrec] = nanim    # animals with records
  sid[nrec] = ks;  did[nrec] = kd
  bi[nrec] = inbb[nanim]
  atbv[nrec] = tbv[nanim]   }
    sel = tbv[1:nanim] + rnorm(nanim,0,SDe*2)
 }
 }
```

Save Data

```
obs = round(obs)  # round off to whole numbers
MYdat = data.frame(anwr,sid,did,year,cgrp,bi,fi,obs,atbv)
MTdat2 = MYdat[1:nrec, ]

save(MYdat2,file="Larry.Rdata")
```

Generate your data.
Retrieve program from website.

Solving MME

- Read in data frame
- Form pedigrees and diagonals of $\mathbf{A}^{-1}$
- Initialize solutions, variance parameters
- Iterate on Data
- Summarize Results

Read Data Frame

```
load(file="Larry.Rdata")
names(MYdat2)
nrec = nrow(MYdat2)
obs = MYdat2$obs; yrfac = MYdat2$year
cgfac = MYdat2$cgrp;    anwr = MYdat2$anwr
bi = MYdat2$bi;       fi=MYdat2$fi
aid = MYdat2$anwr; sid=MYdat2$sid; did=MYdat2$did
atbv = MYdat2$atbv
```

Pedigree Prep

```
mmm = aid[1] - 1
NP = aid[length(aid)]+1
sanm = rep(0,NP); syr=rep(0,NP); scg=rep(0,NP)
aaid = c(1:mmm, aid, NP)
ssid = c(rep(NP,mmm), sid, NP) #all non-zero
ddid = c(rep(NP,mmm), did, NP) #all non-zero
bbi = aaid*0; mwr = mmm + 1
nam = length(aaid) - 1
bbi[1:mmm]=1;  bbi[mwr:nam]=bi;  bbi[NP]=0
```

More Pedigree Prep

```
mtotal = length(aaid)
adiag = rep(0,mtotal)
for(i in 1:nam){
  kis = ssid[i]; kid = ddid[i]
  d = 1/bbi[i];  x = 0.25*d
  adiag[i] = adiag[i] + d
   adiag[kis] = adiag[kis] + x
    adiag[kid] = adiag[kid] + x
  }
  adiag[NP]=0
  zz = c(rep(0,mmm),rep(1,nrec),0)
```

Initialize

```
iters = 0; nmax = 10000; ccc = 100
vare = 64; alpha = 1.6; alphc = 4

while( ccc > 0.000001){
  iters = iters + 1
```

Year Effects - fixed

```
rhs = (obs - sanm[aid] - scg[cgfac])
syr = tapply(rhs,yrfac,mean)  # new solutions
```

CG Effects - random

```
rhs = (obs - sanm[aid] - syr[yrfac])
dcg = tapply(rep(1,nrec),cgfac,sum) + alphac
scg = tapply(rhs,yrfac,sum)/dcg  # new solutions
```

Animal BV - random

```
rhs = (obs - scg[cgfac] - syr[yrfac])
rhs = c( rep(0,mmm),tapply(rhs,aid,sum),0)

ccd = 0; ccs = 0
for(i in nam:1) {
  kis = ssid[i]; kid = ddid[i]; d = alpha/bbi[i]
  x = zz[i] + alpha*adiag[i]
  rhs[i]=rhs[i] + 0.5*d*(sanm[kis]+sanm[kid])
  anew = rhs[i]/x
  diff = anew - sanm[i]
  ccd = ccd + diff*diff
  ccs = ccs + anew*anew
  sanm[i] = anew
  rhs[kis]=rhs[kis]+d*(0.5*sanm[i] - 0.25*sanm[kid])
  rhs[kid]=rhs[kid]+d*(0.5*sanm[i] - 0.25*sanm[kis])
  }
```

End Game

```
 if(iters>nmax)ccc=0
}  # end of while loop

ebv = sanm[aid]  # only for those with records
gtrnd = tapply(ebv,yrfac,mean)
truet = tapply(atbv,yrfac,mean)
plot(gtrnd,type="b", lwd=3, col="red",
  xlab="Year of Birth", ylab="Trait Units")
lines(truet,type="b", lwd=3, col="blue")
title(main="Genetic Change By Year of Birth")
cor(ebv,atbv)
```