Nested or Crossed?

Mila Sun

Winter 2022

https://stats.stackexchange.com/questions/228800/crossed-vs-nested-random-effects-how-do-they-differ-and-how-are-they-specified

```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.6.2
```

```
library(nlme)
```

1 Nested and crossed data

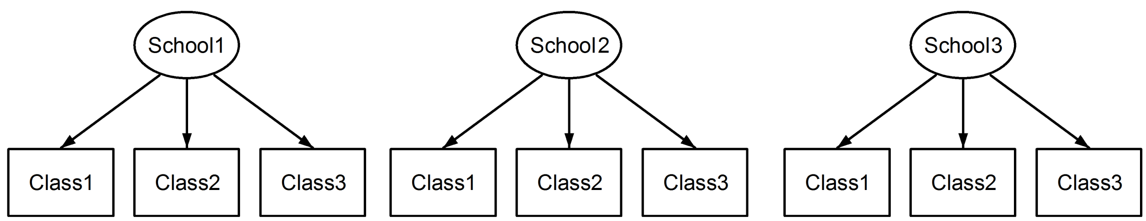


Figure 1: Nested

Here we have classes nested in schools, which is a familiar scenario. The important point here is that, between each school, the classes have the same identifier, even though they are distinct if they are nested. `Class1` appears in `School1`, `School2` and `School3`. However if the data are nested then `Class1` in `School1` is not the same unit of measurement as `Class1` in `School2` and `School3`. If they were the same, then we would have this situation:

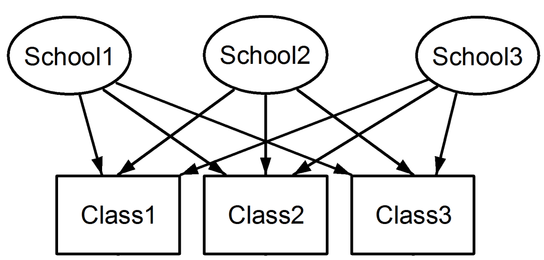


Figure 2: Crossed

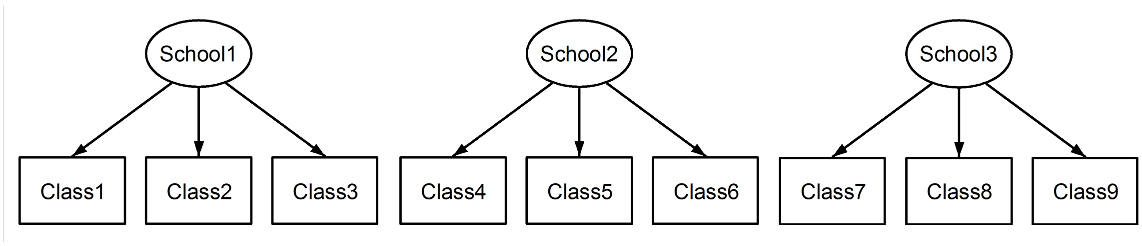
which means that every class belongs to every school. The former is a nested design, and the latter is a crossed design, and we would formulate these in `lme4` using:

- Nested: `(1|School/Class)` or equivalently `(1|School) + (1|Class:School)`
- Crossed: `(1|School) + (1|Class)`

2 Remarks and example

- Nesting is a property of the data, or rather the experimental design, not the model.
- Due to the ambiguity of whether there is nesting or crossing of random effects, it is very important to specify the model correctly as these models will produce different results, as we shall show below.
- Moreover, it is not possible to know, just by inspecting the data, whether we have nested or crossed random effects. This can only be determined with knowledge of the data and the experimental design.

But first let us consider a case where the Class variable is coded uniquely across schools:



There is no longer any ambiguity concerning nesting or crossing. The nesting is explicit. Let us now see this with an example in R, where we have 6 schools (labelled `I-VI`) and 4 classes within each school (labelled `a` to `d`):

```
dt <- read.table("http://bayes.acs.unt.edu:8083/BayesContent/class/Jon/R_SC/Module9/lmm.data.txt",heade
xtabs(~ school + class, dt)
```

```
##       class
## school  a  b  c  d
##    I   50 50 50 50
##    II  50 50 50 50
##    III 50 50 50 50
##    IV  50 50 50 50
##    V   50 50 50 50
##    VI  50 50 50 50
```

We can see from this cross tabulation that every class ID appears in every school, which satisfies your definition of crossed random effects (in this case we have fully, as opposed to partially, crossed random effects, because every class occurs in every school). So this is the same situation that we had in the first figure above. However, if the data are really nested and not crossed, then we need to explicitly tell `lme4`:

```
#a nested model
m0 <- lmer(extro ~ open + agree + social + (1 | school/class), data = dt)
summary(m0)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: extro ~ open + agree + social + (1 | school/class)
##    Data: dt
##
## REML criterion at convergence: 3554.6
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -9.9949 -0.3348  0.0057  0.3394 10.6476
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  class:school (Intercept)  8.2043  2.8643
##  school       (Intercept) 93.8420  9.6872
##  Residual                  0.9684  0.9841
## Number of obs: 1200, groups:  class:school, 24; school, 6
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept) 60.2378227  4.0117903  15.015
## open         0.0061065  0.0049636   1.230
## agree       -0.0076659  0.0056986  -1.345
## social       0.0005404  0.0018524   0.292
##
## Correlation of Fixed Effects:
##        (Intr) open   agree
## open   -0.049
## agree  -0.049 -0.012
## social -0.045 -0.006 -0.009
```

```r
#a crossed model
m1 <- lmer(extro ~ open + agree + social + (1 | school) + (1 |class), data = dt)
summary(m1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: extro ~ open + agree + social + (1 | school) + (1 | class)
##    Data: dt
##
## REML criterion at convergence: 4723.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -7.8677 -0.5421  0.0101  0.5218  8.2282
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  school   (Intercept) 95.888   9.792
##  class    (Intercept)  5.790   2.406
##  Residual              2.787   1.669
## Number of obs: 1200, groups:  school, 6; class, 4
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 60.198841   4.212993  14.289
## open         0.010834   0.008349   1.298
## agree       -0.005420   0.009605  -0.564
## social      -0.001762   0.003107  -0.567
##
## Correlation of Fixed Effects:
##        (Intr) open   agree
## open   -0.079
## agree  -0.079 -0.008
```

```
## social -0.073  0.002 -0.006
```

Now, if we introduce a new variable for the class identifier:

```
dt$classID <- paste(dt$school, dt$class, sep=".")
xtabs(~ school + classID, dt)
```

```
##       classID
## school I.a I.b I.c I.d II.a II.b II.c II.d III.a III.b III.c III.d IV.a IV.b
##    I    50  50  50  50    0    0    0    0     0     0     0     0    0    0
##    II    0   0   0   0   50   50   50   50     0     0     0     0    0    0
##    III   0   0   0   0    0    0    0    0    50    50    50    50    0    0
##    IV    0   0   0   0    0    0    0    0     0     0     0     0   50   50
##    V     0   0   0   0    0    0    0    0     0     0     0     0    0    0
##    VI    0   0   0   0    0    0    0    0     0     0     0     0    0    0
##       classID
## school IV.c IV.d V.a V.b V.c V.d VI.a VI.b VI.c VI.d
##    I      0    0   0   0   0   0    0    0    0    0
##    II     0    0   0   0   0   0    0    0    0    0
##    III    0    0   0   0   0   0    0    0    0    0
##    IV    50   50   0   0   0   0    0    0    0    0
##    V      0    0  50  50  50  50    0    0    0    0
##    VI     0    0   0   0   0   0   50   50   50   50
```

The cross tabulation shows that each level of class occurs only in one level of school. Both model formulations will now produce the same output (that of the nested model `m0` above):

```
m2 <- lmer(extro ~ open + agree + social + (1 | school/classID), data = dt)
summary(m2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: extro ~ open + agree + social + (1 | school/classID)
##    Data: dt
##
## REML criterion at convergence: 3554.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -9.9949 -0.3348  0.0057  0.3394 10.6476
##
## Random effects:
##  Groups         Name        Variance Std.Dev.
##  classID:school (Intercept)  8.2043  2.8643
##  school         (Intercept) 93.8420  9.6872
##  Residual                    0.9684  0.9841
## Number of obs: 1200, groups:  classID:school, 24; school, 6
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept) 60.2378227  4.0117903  15.015
## open         0.0061065  0.0049636   1.230
## agree       -0.0076659  0.0056986  -1.345
## social       0.0005404  0.0018524   0.292
##
## Correlation of Fixed Effects:
##        (Intr) open   agree
```

```
## open   -0.049
## agree  -0.049 -0.012
## social -0.045 -0.006 -0.009
```

3 Summary

- Nested random effects: `(1|group1/group2)` where `group2` is nested within `group1`.
- Crossed random effects are simply: not nested. `(1|group1) + (1|group2)`. Example of crossed effects: fish in lakes.