Gene-Environment Interactions

Sören Möller
moeller@health.sdu.dk

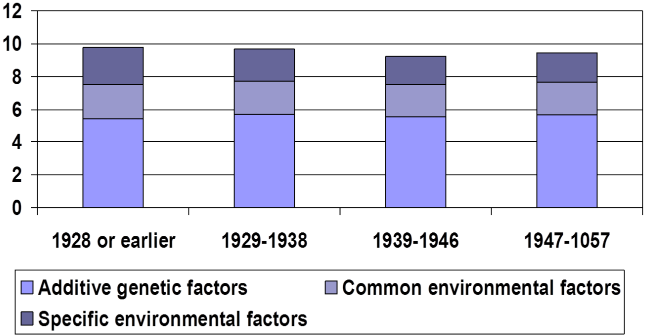
Epidemiology, Biostatistics and Biodemography, Institute of Public Health,
University of Southern Denmark, Denmark

May 2015

What are gene-enviroment interaction

- Until now we assumed that genetic and enviroment effects influence the outcome independently
- It is known that gene-enviroemt (G-E) interactions are very common in other organisms
- Genetic control of sensitivity to enviroment
- Eviromental control of gene expression
- We can not determine the direction of causality between genes and enviroment

Example



The genetic influence on height is different for different birth cohorts in Finland.

[Silventoinen et. al, Am J Publ Health 2000]

G-E interaction in twin models

- Different genetic variation between twins at different levels of exposure
- Example: Rural/urban residence influences genetics of alcohol consumption
- In simple cases, we can stratify the data by enviroment exposure and compare the heritability estimates
- This is only feasible for categorical exposures and large data sets
- And requires same exposure for both twins, to keep the pairs together

Relation to gene-enviroment correlation

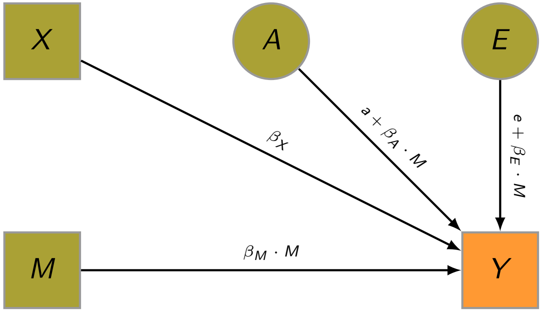
- **GE interaction** genetic control of sensitivity to enviroment
- **GE interaction** eviromental control of gene expression
- Example: Does heritability of IQ depend on socio-econmic status?
- **GE correlation** genetic control of exposure to enviroment
- **GE correlation** enviromental control of gene frequency
- Example: Children with high IQ read more books

Example

There are three possible sources of gene-environment correlation

- Passive gene-environment correlation
    - Parents transmit both their genes and environment
    - Genetically musically talented parents more often listen music and own musical instruments
- Active gene-environment correlation
    - Persons with a certain genotype actively select environments that are correlated with that genotype
    - Genetically musically talented children like to participate musical education
- Reactive gene-environment correlation
    - Persons with a certain genotype evoke certain reactions from environment
    - Music teachers pick up genetically musically talented children for special supervision

Example: univariate GE-path model



$$Y = \beta_X \cdot X + \beta_M M + (a + \beta_a M)A + (e + \beta_e M)E$$
$$Var(Y) = a^2 + (\beta_a M)^2 + e^2 + (\beta_e M)^2$$

The variance of Y is different for different values of M!

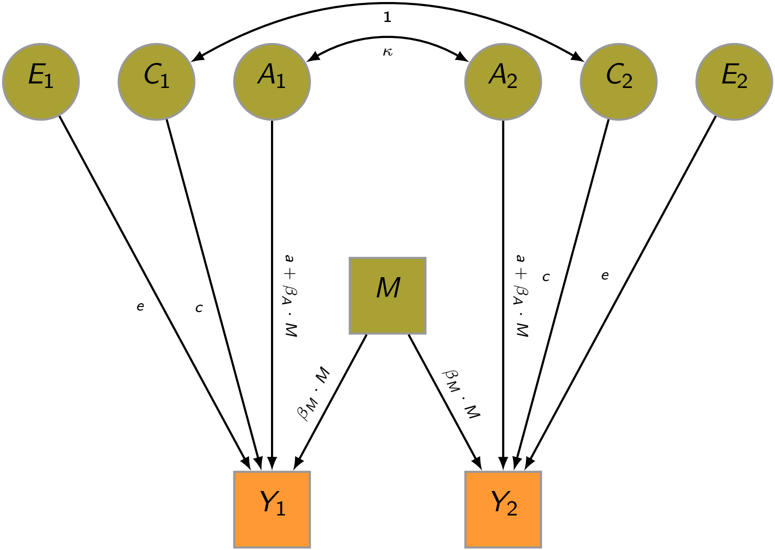
Example: Interpretation

- In this example M not only directly influences Y
- It also **modifies** the strength with which A and E influence Y
- We want to determine β_M, β_A and β_E from our data
- Similar in twin models, where M might be se same or different for the two twins
  - **Common**: Education of parents, place of birth
  - **Individual**: Smoking, income
- Similar interaction with other parts of the biometric models is possible (e.g. C or E)

Comparing with model without interaction

- We want to investigate if there is a gene-enviroment interaction
- No interaction correspons to the corresponding coefficients being 0
- This implies that the model without interaction is a **submodel** of the model with interaction
- Hence we can do log-rank test, to determine if the interaction significantly improves the fit
- In more complex situations, we can compare AIC

Example: ACE with GE-interaction on A



Example: Data

```
head(dwc)
```

```
##        pair  Sex Zyg     bmi1     age1     bmi2     age2
## 1  100001 male  DZ 26.33289 57.57974 25.46939 57.57974
## 4  100004 male  DZ 27.25089 53.51677 28.07504 53.51677
## 6  100006 male  DZ 28.04282 52.57221 22.30936 52.57221
## 7  100007 male  DZ 28.06642 52.49007 26.51180 52.49007
## 9  100011 male  DZ 30.47797 51.35661 27.66010 51.35661
## 11 100013 male  DZ 27.39818 49.89733 25.97012 49.89733
```

How to specify the GE-model: Path coefficients

```
pathA  <- mxMatrix( type="Lower", nrow=1, ncol=1,
  free=TRUE, labels="a11", values=1, lbound=-10, name="a"
pathC  <- mxMatrix( type="Lower", nrow=1, ncol=1,
  free=TRUE, labels="c11", values=1, lbound=-10, name="c"
pathE  <- mxMatrix( type="Lower", nrow=1, ncol=1,
  free=TRUE, labels="e11", values=1, lbound=-10, name="e"
pathAI <- mxMatrix( type="Lower", nrow=1, ncol=1,
  free=TRUE, label="aI11", values=1,  name="aI" )
```

We have added one additional path coefficient, for the effect of M on A.

How to specify the GE-model: Mean structure

```
meanG     <- mxMatrix( type="Full", nrow=1, ncol=1,
  free=TRUE, values=meanBMI, labels="bmi_Mean",
  name="Mean" )
defAge    <- mxMatrix( type="Full", nrow=1, ncol=1,
  free=FALSE, labels=c("data.age1"), name="Age")
pathB     <- mxMatrix( type="Full", nrow=1, ncol=1,
  free=TRUE, values= .01, label="b11", name="b" )
meanAge   <- mxAlgebra( expression= b%*%Age, name="AgeR")
meanGI    <- mxAlgebra( expression=
  cbind((Mean + AgeR),(Mean + AgeR)), name="expMean")
```

We specify a linear effect of M on the mean.
We also have to specify, where the information on the observed values of M comes from. In this case I use age1 for both twins, as it is shared in this dataset.

How to specify the GE-model: Covariance structure

```
covAI     <- mxAlgebra(
  expression=(a+Age%*%aI) %*% t(a+ Age%*%aI), name="AI")
covC      <- mxAlgebra( expression=c %*% t(c), name="C")
covE      <- mxAlgebra( expression=e %*% t(e), name="E")
covMZ     <- mxAlgebra(
  expression= rbind( cbind(AI+C+E , AI+C),
                 cbind(AI+C   , AI+C+E)),
                 name="expCovMZ" )
covDZ     <- mxAlgebra(
  expression= rbind( cbind(AI+C+E, 0.5%x%AI+C),
                 cbind(0.5%x%AI+C ,AI+C+E)),
                 name="expCovDZ" )
```

We add the interaction term to the *A* variance component, and use this instead of the simple component from earlier models.

How to specify the GE-model: Data and expectations

```
dataMZ    <- mxData( observed=mzData, type="raw" )
dataDZ    <- mxData( observed=dzData, type="raw" )

expMZ     <- mxExpectationNormal( covariance="expCovMZ",
  means="expMean", dimnames=c("bmi1","bmi2") )
expDZ     <- mxExpectationNormal( covariance="expCovDZ",
  means="expMean", dimnames=c("bmi1","bmi2") )
funML     <- mxFitFunctionML()
```

This is similar to the ACE model without interaction.

How to specify the GE-model: Running the model

Combining the parts:

```
pars      <- list( pathA, pathC, pathE, covC, covE,
  meanG, pathB, pathAI)
pars2 <- list(pars, funML, covAI, defAge, meanAge, meanGI)
modelMZ   <- mxModel( "MZ", pars2, covMZ, dataMZ, expMZ)
modelDZ   <- mxModel( "DZ", pars2, covDZ, dataDZ, expDZ)
minus2ll  <- mxAlgebra( MZ.objective+ DZ.objective,
  name="minus2sumloglikelihood" )
obj    <- mxFitFunctionAlgebra( "minus2sumloglikelihood" )
ci <- mxCI(c("a11","aI11","c11","e11"))
model  <- mxModel( "Model", pars, modelMZ, modelDZ,
  minus2ll, obj, ci )
```

Running the model:

```
modelFit    <- mxRun(model)
modelFit    <- mxTryHard( modelFit, scale=0.05)
modelSummary    <- summary(modelFit)
```

Results

```
modelSummary


## Summary of Model
##
## The model does not satisfy the first-order optimality conditions to the required accuracy, an
##
## free parameters:
##       name matrix row col     Estimate   Std.Error lbound ubound
## 1      a11      a   1   1 2.134793e+00 0.211171566    -10
## 2      c11      c   1   1 5.210579e-05 0.517328290    -10
## 3      e11      e   1   1 2.006712e+00 0.034256598    -10
## 4 bmi_Mean   Mean   1   1 1.885256e+01 0.261431958
## 5      b11      b   1   1 1.264819e-01 0.005815820
## 6     aI11     aI   1   1 1.531942e-02 0.004666285
##
## confidence intervals:
##             lbound     estimate     ubound note
## a11   1.709567114 2.134793e+00 2.55047417
## aI11  0.006206473 1.531942e-02 0.01647262
## c11  -0.752967446 5.210579e-05 0.75296643
## e11   1.941007884 2.006712e+00 2.07547452
##
## observed statistics:  8542
## estimated parameters:  6
## degrees of freedom:  8536
## fit value ( -2lnL units ):  44266
## number of observations:  4271
## Information Criteria:
##       |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
## AIC:       27194.00               44278.00                       NA
## BIC:      -27091.57               44316.16                  44297.1
```

Testing for interaction

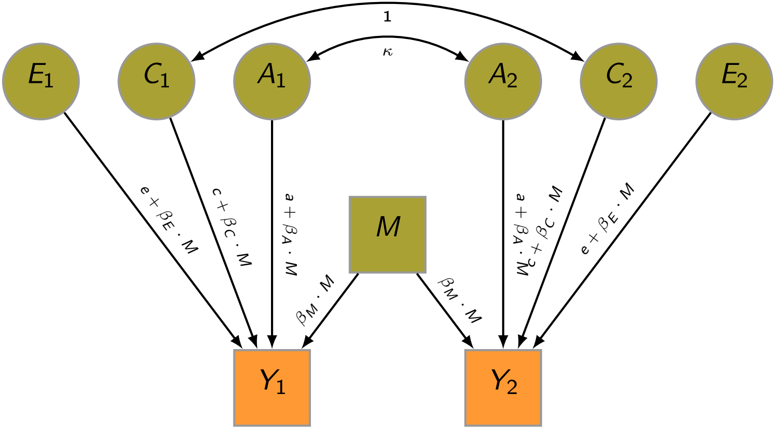
We can modify parts of our model, to compare slightly different models:

```
modelACE    <- mxModel( model, name="modelACE" )
modelACE    <- omxSetParameters( modelACE, label="aI11", free=F, values=0)

modelACEFit    <- mxRun(modelACE)
modelACEFit    <- mxTryHard(modelACEFit, scale=0.05)
modelACESummary    <- summary(modelACEFit)
```

We then can test for a significant difference between the models:

```
mxCompare(modelFit, modelACEFit)


##    base comparison ep minus2LL   df      AIC   diffLL diffdf            p
## 1 Model       <NA>  6 44266.00 8536 27194.00       NA     NA           NA
## 2 Model   modelACE  5 44276.91 8537 27202.91 10.90997      1 0.0009564808
```

Excercise



Fit an ACE model on the twinbmi data, which has interaction of age on all the (A, C, E) components. And test if this model fits better than the example above.

Solution: Specification

```
pathAI    <- mxMatrix( type="Full", nrow=1, ncol=1,
free=TRUE, values=.6, label="aI11", name="aI" )
pathCI    <- mxMatrix( type="Full", nrow=1, ncol=1,
free=TRUE, values=0, label="cI11", name="cI" )
pathEI    <- mxMatrix( type="Full", nrow=1, ncol=1,
free=TRUE, values=0, label="eI11", name="eI" )
covAI     <- mxAlgebra(
expression=(a+ Age%*%aI) %*% t(a+ Age%*%aI), name="AI" )
covCI     <- mxAlgebra(
expression=(c+ Age%*%cI) %*% t(c+ Age%*%cI), name="CI" )
covEI     <- mxAlgebra(
expression=(e+ Age%*%eI) %*% t(e+ Age%*%eI), name="EI" )
covMZ     <- mxAlgebra(
expression= rbind( cbind(AI+CI+EI , AI+CI),
cbind(AI+CI   , AI+CI+EI)), name="expCovMZ" )
covDZ     <- mxAlgebra(
expression= rbind( cbind(AI+CI+EI, 0.5%x%AI+CI),
cbind(0.5%x%AI+CI ,AI+CI+EI)), name="expCovDZ" )
```

Solution: Specification

```
pars      <- list( pathA, pathC, pathE, meanG, pathB,
  pathAI, pathCI, pathEI)
pars2     <- list(covAI, covCI, covEI,  defAge, meanAge,
  meanGI)
modelMZ   <- mxModel( "MZ", pars, pars2, covMZ, dataMZ,
  expMZ, funML)
modelDZ   <- mxModel( "DZ", pars, pars2, covDZ, dataDZ,
  expDZ, funML)
minus2ll  <- mxAlgebra( MZ.objective+ DZ.objective,
name="minus2sumloglikelihood" )
obj       <- mxFitFunctionAlgebra( "minus2sumloglikelihood
ci <- mxCI(c("a11","aI11","c11","cI11","e11","eI11"))
model  <- mxModel( "Model", pars, modelMZ, modelDZ,
  minus2ll, obj, ci )

modelFit    <- mxRun(model)
modelFit    <- mxTryHard( modelFit, scale=0.05)
modelSummary    <- summary(modelFit)
```

Solution: Results

```
modelSummary
```

```
## Summary of Model
##
## The model does not satisfy the first-order optimality conditions to the required accuracy, an
##
## free parameters:
##       name matrix row col       Estimate   Std.Error lbound ubound
## 1      a11      a   1   1 -2.504510e+00 0.248746138    -10
## 2      c11      c   1   1  5.223779e-04          NA    -10
## 3      e11      e   1   1 -1.429516e+00 0.204380567    -10
## 4 bmi_Mean   Mean   1   1  1.886522e+01 0.261286360
## 5      b11      b   1   1  1.261991e-01 0.005801291
## 6     aI11     aI   1   1 -6.943577e-03 0.005593962
## 7     cI11     cI   1   1 -1.012522e-05          NA
## 8     eI11     eI   1   1 -1.299820e-02 0.004610746
##
## confidence intervals:
##            lbound      estimate      ubound note
## a11  -2.995543516 -2.504510e+00 -1.75716157
## aI11 -0.007364664 -6.943577e-03  0.00428740
## c11  -2.993916600  5.223779e-04  2.99391567
## cI11 -0.058651739 -1.012522e-05  0.05865160
## e11  -1.838763039 -1.429516e+00 -1.02861546
## eI11 -0.022109857 -1.299820e-02 -0.01194523
##
## observed statistics:  8542
## estimated parameters:  8
## degrees of freedom:  8534
## fit value ( -2lnL units ):  44257.85
## number of observations:  4271
```

Solution: Test

```
modelACEGA    <- mxModel( model, name="modelACEGA" )
modelACEGA    <- omxSetParameters( modelACEGA,
  label="cI11", free=F, values=0)
modelACEGA    <- omxSetParameters( modelACEGA,
  label="eI11", free=F, values=0)

modelACEGAFit    <- mxRun(modelACEGA)
modelACEGAFit    <- mxTryHard(modelACEGAFit, scale=0.05)
modelACEGASummary    <- summary(modelACEGAFit)
```

```
mxCompare(modelFit, modelACEGAFit)


##    base comparison ep minus2LL   df      AIC   diffLL diffdf          p
## 1 Model       <NA>  8 44257.85 8534 27189.85       NA     NA         NA
## 2 Model modelACEGA  6 44266.00 8536 27194.00 8.154919      2 0.01695048
```