# Mixed model with LMMstar - Part 3

## Estimation, statistical inference, and prediction

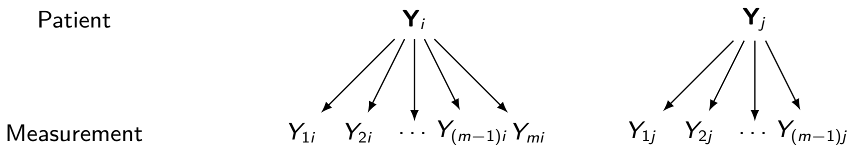
**Brice Ozenne**[1,2], Julie Lyng Forman[2]
brice.ozenne@nru.dk

[1] Section of Biostatistics, Department of Public Health, University of Copenhagen.
[2] Neurobiology Research Unit, University Hospital of Copenhagen, Rigshospitalet.

June 7th 2022, Method week at Karolinska Institutet

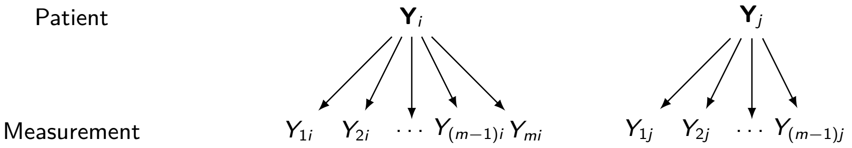
# Recap'

**Hierarchical data**: $p$ (baseline) covariates $\mathbf{X}$ and $m$ outcomes $\mathbf{Y}$



# Recap'

**Hierarchical data**: $p$ (baseline) covariates $\mathbf{X}$ and $m$ outcomes $\mathbf{Y}$



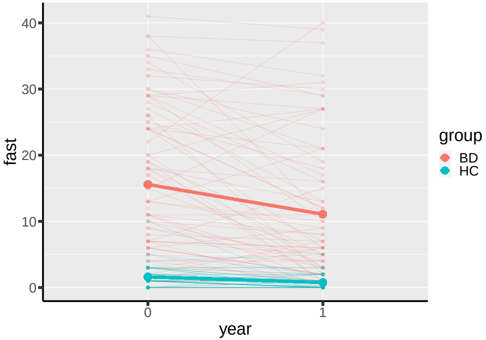

**Statistical model**: multivariate Gaussian

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \boldsymbol{\varepsilon}_i, \text{ where } \boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \dots, \varepsilon_{mi}) \sim \mathcal{N}(0, \Omega(\mathbf{X}_i, \gamma))$$

- $\mu(\mathbf{X}, \beta) = \mathbf{X}\beta$ modeled mean
- $\Omega(\mathbf{X}, \gamma)$ modeled residual variance-covariance.
- $\Theta = (\beta, \gamma)$ model parameters

# Example: abeta study

```
eLMM.abeta <- lmm(fast ~ group*year,
    repetition = ~year|id,
    structure = UN(~group),
    data = abetaL, control = list(optimizer = "FS"))
```



# Notation: abeta study

$\widehat{\beta}$

```
coef(eLMM.abeta, effects = "mean")
```

```
(Intercept)      groupHC       year1 groupHC:year1
  15.574713   -13.961076   -4.489151      3.642689
```

$\widehat{\gamma}$

```
coef(eLMM.abeta, effects = c("var","cor"))
```

```
  sigma:BD    sigma:HC      k.1:BD      k.1:HC rho(0,1):BD rho(0
11.5008923   1.7812448   0.9468697   0.6331572   0.7531450    0.3
```

$\widehat{\Omega}$

```
sigma(eLMM.abeta)
```

```
$BD                          $HC
         0        1                    0         1
0 132.2705  94.3261          0 3.1728327 0.6096141
1  94.3261 118.5888          1 0.6096141 1.2719508
```

## Part 3

How does the software estimates $\Theta = (\beta, \gamma)$?
How can the user tune the corresponding optimization procedure?

- argument `control` in `lmm`

How can we contrast the estimates $\widehat{\Theta} = (\widehat{\beta}, \widehat{\gamma})$?

- method `anova` (linear contrasts)
  and `estimate` (non-linear contrasts)

How can we predict new/missing outcome values,
e.g. get $\widehat{\mathbb{E}}[Y_{im}|\mathbf{X}_i]$?
$\widehat{\mathbb{E}}[Y_{im}|\mathbf{X}_i, Y_{1i}, \ldots, Y_{(m-1)i}]$?

- method `prediction`

# Estimation

# General idea

Model parameters $\Theta = (\beta, \gamma)$ are estimated by maximizing an **objective function** with respect to the data

- what parameters makes the observed data most likely?

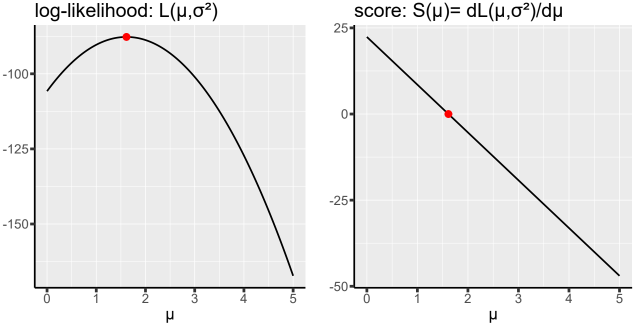
# General idea

Model parameters $\Theta = (\beta, \gamma)$ are estimated by maximizing an **objective function** with respect to the data

- what parameters makes the observed data most likely?

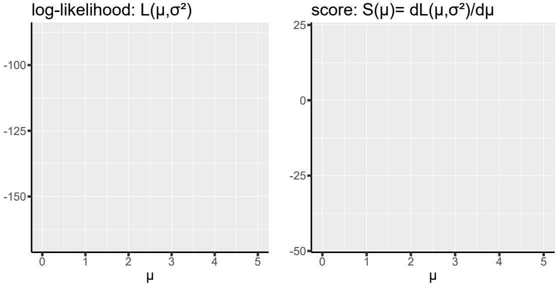
*Toy example*: univariate linear model

$$Y_1 = \mu + \varepsilon \text{ where } \varepsilon \sim \mathcal{N}\left(0, \sigma^2\right)$$



# Gradient descent

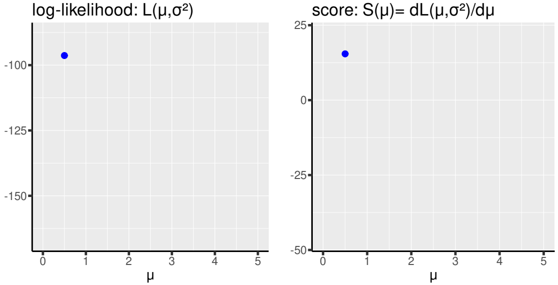
Idea: find the maximum, i.e. when the score is 0



# Gradient descent

Idea: find the maximum, i.e. when the score is 0

- start somewhere: $\mu_0$

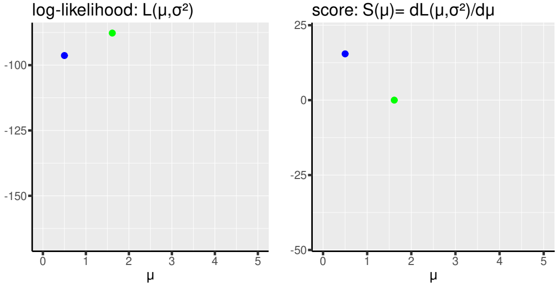


# Gradient descent

Idea: find the maximum, i.e. when the score is 0

- start somewhere: $\mu_0$
- learn from mistake: evaluate the "error" $\mathcal{S}(\mu_0)$
  update to reduce the error

$$\mu_1 = \mu_0 + \alpha \mathcal{S}(\mu_0), \text{ often } \alpha = \mathcal{I}^{-1}(\mu_0)$$



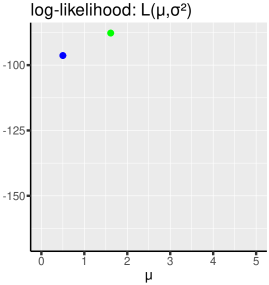
# Gradient descent

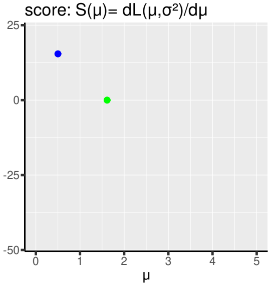
Idea: find the maximum, i.e. when the score is 0

- start somewhere: $\mu_0$
- learn from mistake: evaluate the "error" $\mathcal{S}(\mu_0)$
  update to reduce the error

$$\mu_1 = \mu_0 + \alpha\mathcal{S}(\mu_0),\ \text{often}\ \alpha = \mathcal{I}^{-1}(\mu_0)$$

- iterate until the error is neglectable, i.e. $\mathcal{S}(\mu_k) \approx 0$



# FS optimizer

Initialization:

- ordinary least square for $\beta$
- residual variance/correlation for $\gamma$

Estimation by iterating between:

- gradient descent for $\gamma$: $\gamma_{k+1} = \gamma_k + \alpha \mathcal{S}(\gamma_k)$ (given $\beta$)
- compute the GLS estimator of $\beta$ (given $\gamma$)
- check convergence and increase in log-likelihood[1]

[1] otherwise partial update (e.g. $\alpha \leftarrow \alpha/2$)

# FS optimizer

Initialization:

- ordinary least square for $\beta$
- residual variance/correlation for $\gamma$

Estimation by iterating between:

- gradient descent for $\gamma$: $\gamma_{k+1} = \gamma_k + \alpha \mathcal{S}(\gamma_k)$ (given $\beta$)
- compute the GLS estimator of $\beta$ (given $\gamma$)
- check convergence and increase in log-likelihood[1]

⚠ to satisfy constraints on $\gamma$, e.g. variance must be positive

- log-transform for variance, atanh transform for correlation

[1] otherwise partial update (e.g. $\alpha \leftarrow \alpha/2$)

# FS in practice

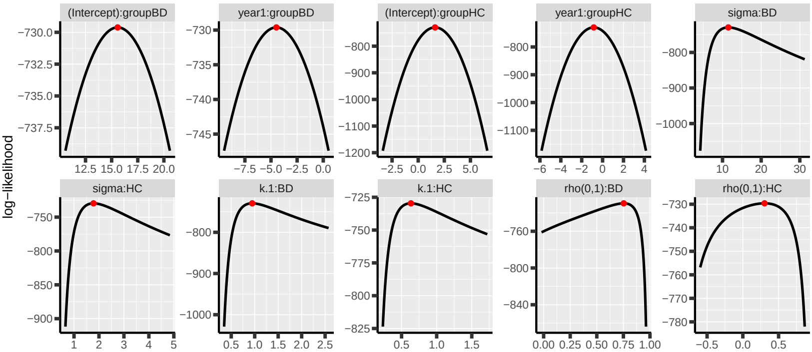
```
eLMM.abeta_fs <- lmm(fast ~ group*year, data = abetaL,
       repetition = ~year|id, structure = UN(~group),
       control = list(optimizer = "FS", trace = 5))
```

```
Initialization:
   (Intercept)       groupHC         year1 groupHC:year1       si
    15.5747126   -13.9610763    -4.5473154     3.7141668     11.5
Loop:
 iteration 1: logLik=-729.624041
           (Intercept)       groupHC        year1 groupHC:year1
 estimate 1.557471e+01 -1.396108e+01 -4.48915269    3.64270819
 diff     1.207923e-13 -1.207923e-13  0.05816269   -0.07145864
 score              NA            NA          NA            NA
              sigma:BD     sigma:HC      k.1:BD        k.1:HC
 estimate 1.150101e+01 1.781274e+00  0.946870688  0.6331082708
 diff     1.163441e-04 2.879615e-05 -0.003847878 -0.0009589621
 score    9.781768e-03 7.494316e-03 -0.684912654 -0.0648472211
```

# Visualization of the objective function

"FS" performs Restricted Maximum Likelihood (REML), i.e. maximizes:

$$\mathcal{L}(\Theta|\mathbf{Y},\mathbf{X}) \propto -\frac{1}{2}\log\left|\sum_{i=1}^{n}\mathbf{X}_i\Omega^{-1}(\mathbf{X}_i,\gamma)\mathbf{X}_i^{\mathsf{T}}\right| - \frac{1}{2}\sum_{i=1}^{n}\log|\Omega(\mathbf{X}_i,\gamma)| + (\mathbf{Y}_i-\mathbf{X}_i\beta)\Omega(\mathbf{X}_i,\gamma)^{-1}(\mathbf{Y}_i-\mathbf{X}_i\beta)^{\mathsf{T}}$$



# Other optimizers

gls:

✔ parameter transformation ensuring positive definite $\Omega$ (Pinheiro and Bates, 1996)

✘ no feedback from the optimization procedure

optim optimizer (e.g. BFGS):

- simulateneous search over $\beta$ and $\gamma$ (do not use GLS estimator)

# Exercise 1

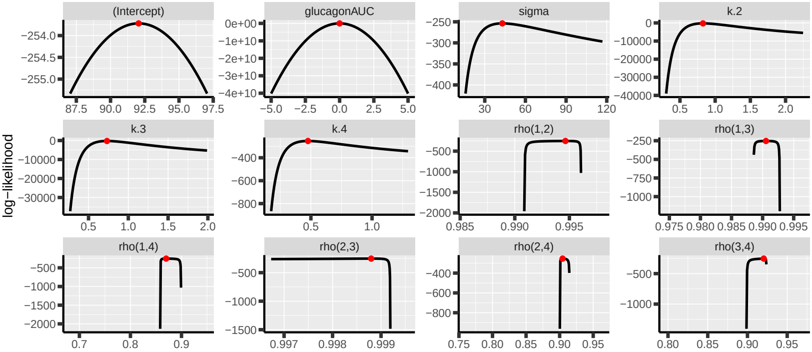
The following code returns a convergence issue:

```
eLMM.debug <- lmm(weight ~ glucagonAUC,
    data = gastricbypassL,
    repetition = ~visit|id,
    structure = "UN",
    control = list(optimizer = "FS")
    ) # line 108
```

- what is the issue?
- can you find remedies?
  - choice of the optimizer
  - number of iterations
  - starting point
  - ...

# Visualization of the log-likelihood

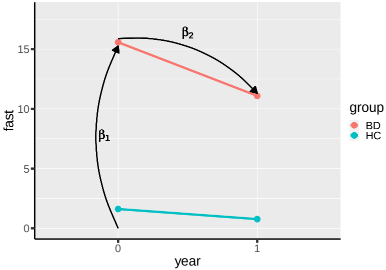
Ill-defined model due to lack of `visit` effects on the mean



# Statistical inference

# Contrasting estimates

The optimization procedure provides estimates, e.g. $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$:

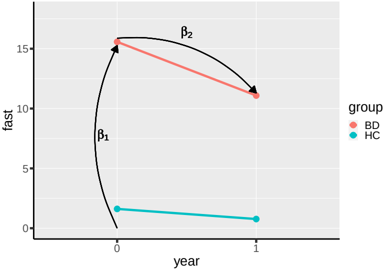


# Contrasting estimates

The optimization procedure provides estimates, e.g. $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$:

But we may be interested in different combination of coefficients:

- time evolution
  $\widehat{\psi}_1 = \widehat{\beta}_2$
- value at year 1
  $=$ baseline value
  $+$ time evolution
  $\widehat{\psi}_2 = \widehat{\beta}_1 + \widehat{\beta}_2$

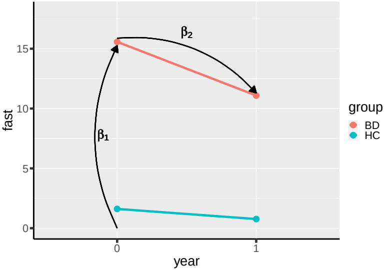


# Contrasting estimates

The optimization procedure provides estimates, e.g. $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)$:

But we may be interested in different combination of coefficients:

- time evolution
  $\widehat{\psi}_1 = \widehat{\beta}_2$
- value at year 1
  $=$ baseline value
  $+$ time evolution
  $\widehat{\psi}_2 = \widehat{\beta}_1 + \widehat{\beta}_2$




$$\widehat{\psi} = C\widehat{\beta} \text{ where } C = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \text{ (contrast matrix)}$$

# Contrasting estimates with anova

Using a "text" formula:

```
anova(eLMM.abeta,
      effects = c("groupHC=0",
    "groupHC+groupHC:year1=0"))
```

Can be named:

```
eANOVA.group <- anova(eLMM.abeta,
        effects = c("baseline" = "groupHC=0",
      "follow-up" = "groupHC+groupHC:year1=0"))
```

# Contrasting estimates with anova

Using a "text" formula:

```
anova(eLMM.abeta,
      effects = c("groupHC=0",
    "groupHC+groupHC:year1=0"))
```

Can be named:

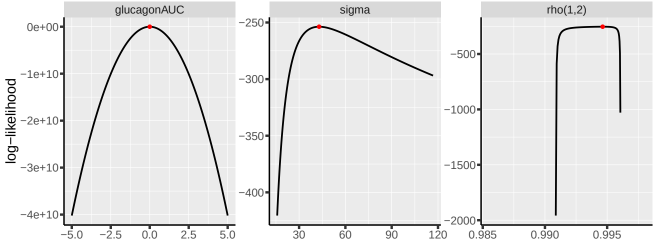
```
eANOVA.group <- anova(eLMM.abeta,
        effects = c("baseline" = "groupHC=0",
      "follow-up" = "groupHC+groupHC:year1=0"))
```

⚠ should refer to model parameters not variables
(e.g. groupHC not "group")

# Quantifying uncertainty

The optimization procedure quantifies the precision of $\widehat{\beta}$

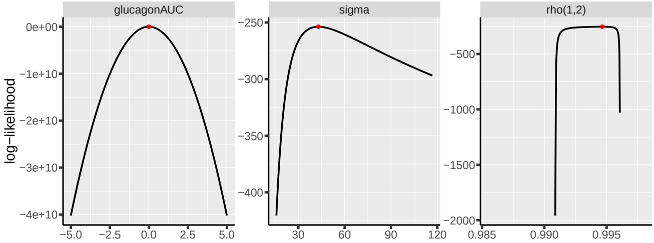
- information matrix $\widehat{\mathcal{I}}_{\widehat{\Theta}} = -\frac{\partial^2 \mathcal{L}(\widehat{\Theta}|\mathcal{O})}{\partial \Theta^2}$
  $\rightarrow$ curvature of the log-likelihood



# Quantifying uncertainty

The optimization procedure quantifies the precision of $\widehat{\beta}$

- information matrix $\widehat{\mathcal{I}}_{\widehat{\Theta}} = -\frac{\partial^2 \mathcal{L}(\widehat{\Theta}|\mathcal{O})}{\partial \Theta^2}$
  $\rightarrow$ curvature of the log-likelihood




The variance of a linear contrast can be deduced:

- $\widehat{\Sigma}_{\widehat{\psi}} = C\widehat{\mathcal{I}}_{\widehat{\beta}}^{-1}C^{\intercal}$

# Statistical tests

Univariate Wald test:

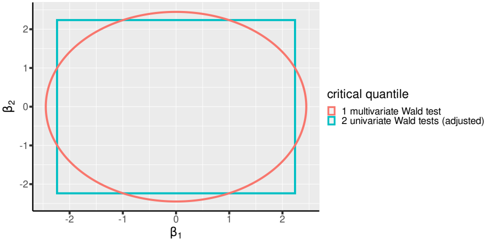
- compare the estimate to its standard error

$$\frac{\widehat{\beta}_1}{\sqrt{\Sigma_{\widehat{\beta}_1}}}$$

Multivariate Wald test (F-test):

- distance to the orgine of the vector of estimates (after normalization by the uncertainty)

$$\widehat{\beta}\Sigma_{\widehat{\beta}}^{-1}\widehat{\beta}^{\intercal}$$



# Wald/F-tests

```
summary(eANOVA.group, method = "none")
```

```
        || User-specified linear hypotheses ||

 - Multivariate Wald test (global null hypothesis)
 F-statistic df.num df.denom p.value
      61.517      2   88.908       0 ***

 - Univariate Wald test (individual null hypotheses)
          estimate       se       df     lower   upper p.value
baseline  -13.9611   1.2619  93.9457 -16.4667 -11.456 < 1e-05 ***
follow-up -10.3184   1.2277  86.2881 -12.7588  -7.878 < 1e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Standard errors: model-based
(CIs/p-values not adjusted for multiple comparisons)
```

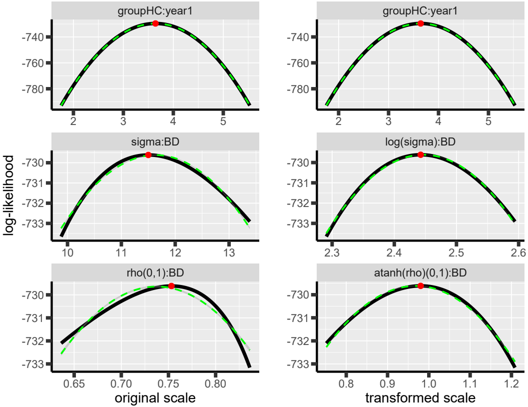
# Distributional results

Estimates are asymptotically jointly normally distributed and so are linear contrasts $C\widehat{\beta}$.

# Distributional results

Estimates are asymptotically jointly normally distributed and so are linear contrasts $C\widehat{\beta}$.

True in finite sample when the log-likelihood is (locally) quadratic (Geyer, 2013)



# Adjustment for multiple comparisons

```
summary(eANOVA.group)
```

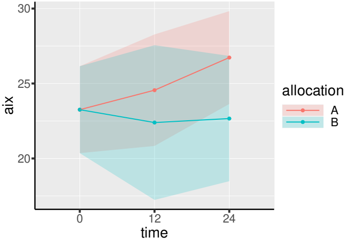
```
        || User-specified linear hypotheses ||

 - Multivariate Wald test (global null hypothesis)
 F-statistic df.num df.denom p.value
      61.517      2   88.908       0 ***

 - Univariate Wald test (individual null hypotheses)
          estimate       se       df    lower    upper p.value
baseline  -13.9611   1.2619  93.9457 -16.7612 -11.1609 < 1e-05 ***
follow-up -10.3184   1.2277  86.2881 -13.0425  -7.5943 < 1e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Standard errors: model-based
(CIs/p-values adjusted for multiple comparisons -- max-test adjust
Adjusted CIs/p-values computed using 1e+05 samples.
```

# Exercise 2

```
lmm(aix ~ time + time:treat,
    repetition = ~time:treat|id,
    structure = UN,
    data = ckdL,
    control = list(optimizer = "FS")
    ) # line 207
```




```
(Intercept)    time12    time24 time12:treatB time24:treatB
   23.25499 1.299417 3.475746     -2.158075     -4.069283
```

Compute the average time effect in group B

- how much patients differ from baseline at 12 and 24 weeks?

# Predictions

# Static vs. dynamic predictions

**Static**: prediction conditional on the covariates:

- $\mathbb{E}[Y_{1i}|X_i] = X_{1i}\beta$
  $\mathbb{E}[Y_{mi}|X_i] = X_{mi}\beta$

**Dynamic**: prediction conditional on the covariates and observed outcomes from the same cluster:

$$\mathbb{E}[Y_{im}|X_i, Y_{1i}] = \mathbb{E}[Y_{im}|X_i] + \mathbb{C}ov(Y_{mi}, Y_{1i})\frac{Y_{1i} - \mathbb{E}[Y_{1i}|X_i]}{\mathbb{V}ar[Y_{1i}]}$$

# Static vs. dynamic predictions

**Static**: prediction conditional on the covariates:

- $\mathbb{E}[Y_{1i}|X_i] = X_{1i}\beta$
  $\mathbb{E}[Y_{mi}|X_i] = X_{mi}\beta$

**Dynamic**: prediction conditional on the covariates and observed outcomes from the same cluster:

$$\mathbb{E}[Y_{im}|X_i, Y_{1i}] = \mathbb{E}[Y_{im}|X_i] + \mathbb{C}ov(Y_{mi}, Y_{1i}) \frac{Y_{1i} - \mathbb{E}[Y_{1i}|X_i]}{\mathbb{V}ar[Y_{1i}]}$$

$$= X_{mi}\beta + \underbrace{\rho_i(m,1)\,\sigma_{mi} \underbrace{\frac{Y_{1i} - X_{1i}\beta}{\sigma_{1i}}}_{\text{normalized residual}}}_{\text{rescaled on } Y_m}$$

# Prediction - illustration (1/2)

Consider the dataset:

```
dfi <- abetaL[abetaL$id==2,
       c("id","group","year","fast")]
dfi
```

```
  id group year fast
3  2    BD    0   32
4  2    BD    1   NA
```

and the mixed model:

```
eLMM.abeta <- lmm(fast ~ 0+group+group:year,
    repetition = ~year|id, structure = UN(~group),
    data = abetaL, control = list(optimizer = "FS"))
eTHETA <- coef(eLMM.abeta, effects = "all")
```

```
groupBD:year1 groupHC:year1
   -4.4891511    -0.8464617
```

# Prediction - illustration (2/2)

Static prediction:

```
predict(eLMM.abeta, newdata = dfi[2,,drop=FALSE])
```

```
  estimate       se      df   lower    upper
1 11.08556 1.215013 82.8635 8.66889 13.50223
```

Normalized residual:

```
(dfi$fast[1] - eTHETA["groupBD"])/eTHETA["sigma:BD"]
```

```
 groupBD
1.428175
```

Dynamic prediction:

```
predict(eLMM.abeta, newdata = dfi,
 type = "dynamic", keep.newdata = TRUE)
```

```
  id group year fast estimate       se  df    lower    upper
1  2    BD    0   32       NA       NA  NA       NA       NA
2  2    BD    1   NA 22.79893 2.215132 Inf 18.45735 27.14051
```

# Imputation - illustration

Combined observed and predicted outcome values:

```
dfi <- fitted(eLMM.abeta, impute = TRUE,
      format = "wide")
dfi[1:4,]
```

```
  id fast.0 imputed.0   fast.1 imputed.1
1  1      1     FALSE  0.00000     FALSE
3  2     32     FALSE 22.79893      TRUE
5  3     29     FALSE 31.00000     FALSE
7  4      1     FALSE  6.00000     FALSE
```

Average within group:

```
abetaA <- merge(dfi, abetaW, by = "id", sort = FALSE)
tapply(abetaA$fast.1-abetaA$fast.0, abetaA$group, mean)
```

```
        BD         HC
-4.4891511 -0.8464617
```

# Exercise 3

Consider the following model:

```
eLMM.ckd <- lmm(aix ~ 0 + time + time:treat,
  repetition = ~time.treat|id, structure = UN,
  data = ckdL, control = list(optimizer = "FS")
  ) # line 346
```

and dataset:

```
ckdL[ckdL$id==29,c(1,6,8,10,11)]
```

```
    id time  aix treat time.treat
1   29    0 15.5     A        0.A
52  29   12 18.0     B       12.B
103 29   24   NA     B       24.B
```

What is the expected aix value for individual 29 at time 24:

- given its group membership and time?
- given its group membership, time, and the observed aix?

# Conclusion

# Conclusion

LMMstar implements "covariance pattern" mixed models, with focus on:

- user-friendly interface
- statistical inference
- model diagnostic

More details in the vignette of the package.

Possible improvements:

- more covariance patterns
- more robust optimizer (e.g. spherical reparametrization)
- computational speed

# Bug report

Unexpected behavior of the LMMstar package can be reported to:
`https://github.com/bozenne/LMMstar/issues`.
Note: this makes your request visible to other users,

Please include a **minimal reproducible example** [2] in your report, otherwise it is likely that we will not be able to identify and solve the issue.

We do NOT provide support for other packages or R programming in general.

[2] `https://stackoverflow.com/help/minimal-reproducible-example`

# Comments, suggestions?

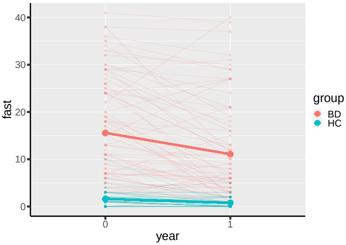
# Reference I

Geyer, C. J. (2013). Asymptotics of maximum likelihood without the lln or clt or sample size going to infinity. In *Advances in Modern Statistical Theory and Applications: A Festschrift in honor of Morris L. Eaton*, pages 1–24. Institute of Mathematical Statistics.

Pinheiro, J. C. and Bates, D. M. (1996). Unconstrained parametrizations for variance-covariance matrices. *Statistics and computing*, 6(3):289–296.

# Exercise 2 bis

```
lmm(fast ~ group*year,
    repetition = ~year|id,
    structure = UN(~group),
    data = abetaL,
    control = list(optimizer = "FS")
    ) # line 243
```




Investigate whether:

- the variance structure depends on time and disease status.
- the correlation structure depends on disease status.

⚠ the scale used for statistical inference matters (e.g. original vs. log/atanh), see arguments `transform.sigma` and `transform.rho`.