Non-continuous data examples

Emmanuelle, Belhal

31/03/2020

Context

Evaluate the performance of saemix in non-continuous data models, and debug if necessary.

Objective

- check the algorithm on examples from the monolix documentation
  - for each type of model
  - comparing the results to Monolix run through the lixoftconnector library
- set up simulations to assess estimation of population parameters for non-continuous data models
  - binary data: SIR example Marilou
  - categorical data: example in testbelhal
  - count data: example in testbelhal
  - TTE: example in testbelhal
  - RTTE: documentation example from demo.R
- debug why example from Ana isn't working

Notes

1) for ORD data model, the response is a predictor. Test with new data without individual observations is non applicable.
2) For ORD data: problem in estimating parameters with new data (map and pop params) NEED TO DEBUG. Could be in map.saemix???
3) COUNT data model: WHEN ONLY ONE PARAM TO ESTIMATE (fixed.estim=c(1,0)) OBTAIN:

```
# Error in cbind(blocA, t(blocC)) :
#   le nombre de lignes des matrices doit correspondre (voir argument 2)
```

Simulation settings

Set up initial conditions fairly standard as in the original examples.

In a second step, investigate impact of eg number of subjects, number of samples/design, IIV to assess performance more fully. Maybe couple these simulations with investigation of approaches to estimate SE.

Methods to estimate SE

- bootstrap methods
  - including conditional bootstrap
- new SE approach to be developed by Melanie
- SIR

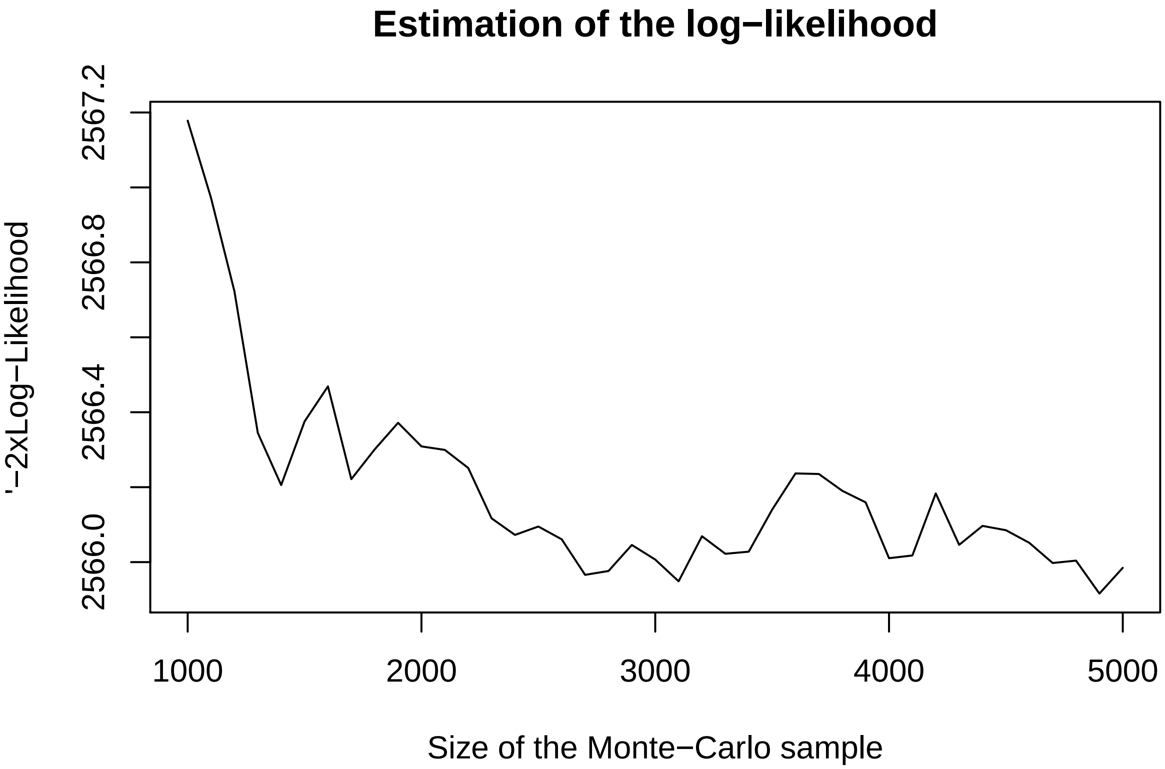
1. Test examples from the Monolix documentation

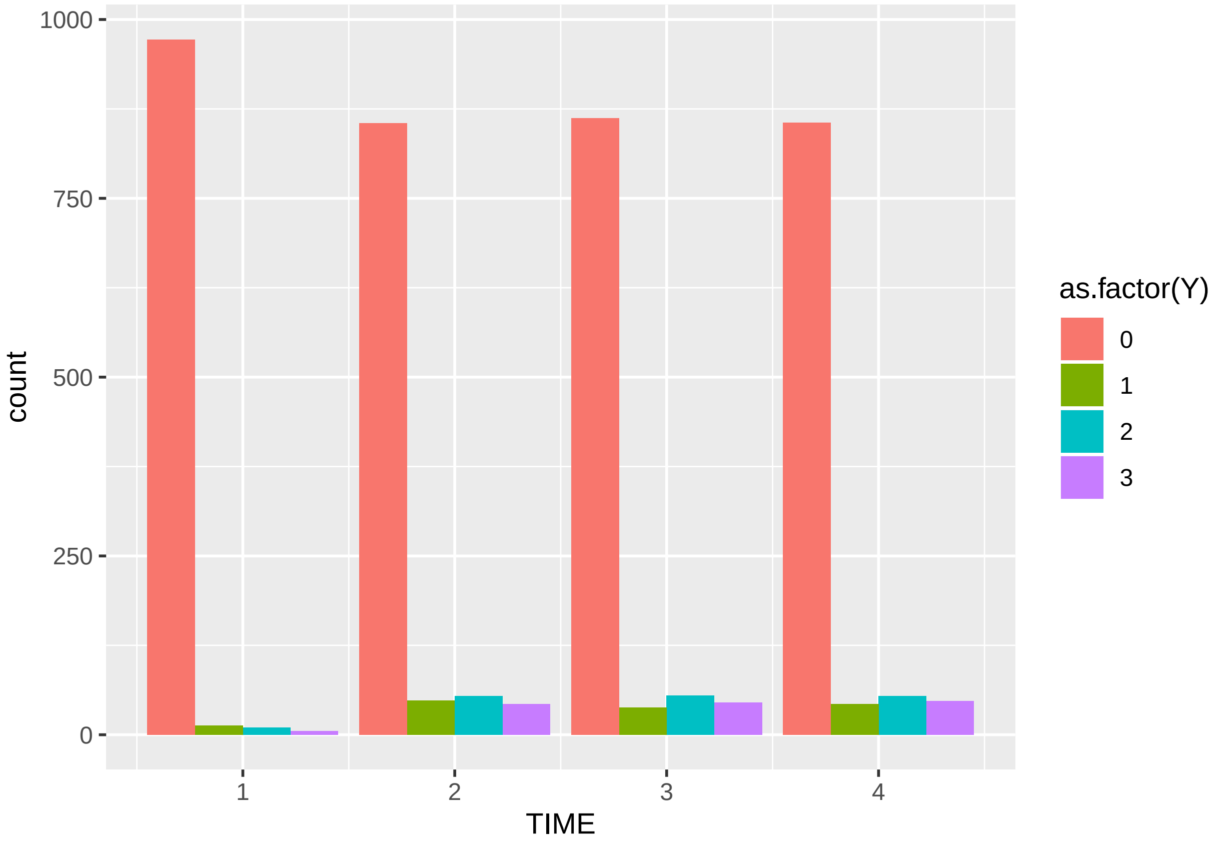
Ordinal data

Modified from initial example to introduce a period effect. Corresponds to exemple *categorical2_data.mlxtran* without a dose effect (adding a dose effect doesn't work in saemix).

Notes

- lots of error messages about the optimisation step in this example because only one random effect, look into it !
- also error reading the data (complains about a missing column but still manages to produce the data object)
- check why the model with a dose effect doesn't work



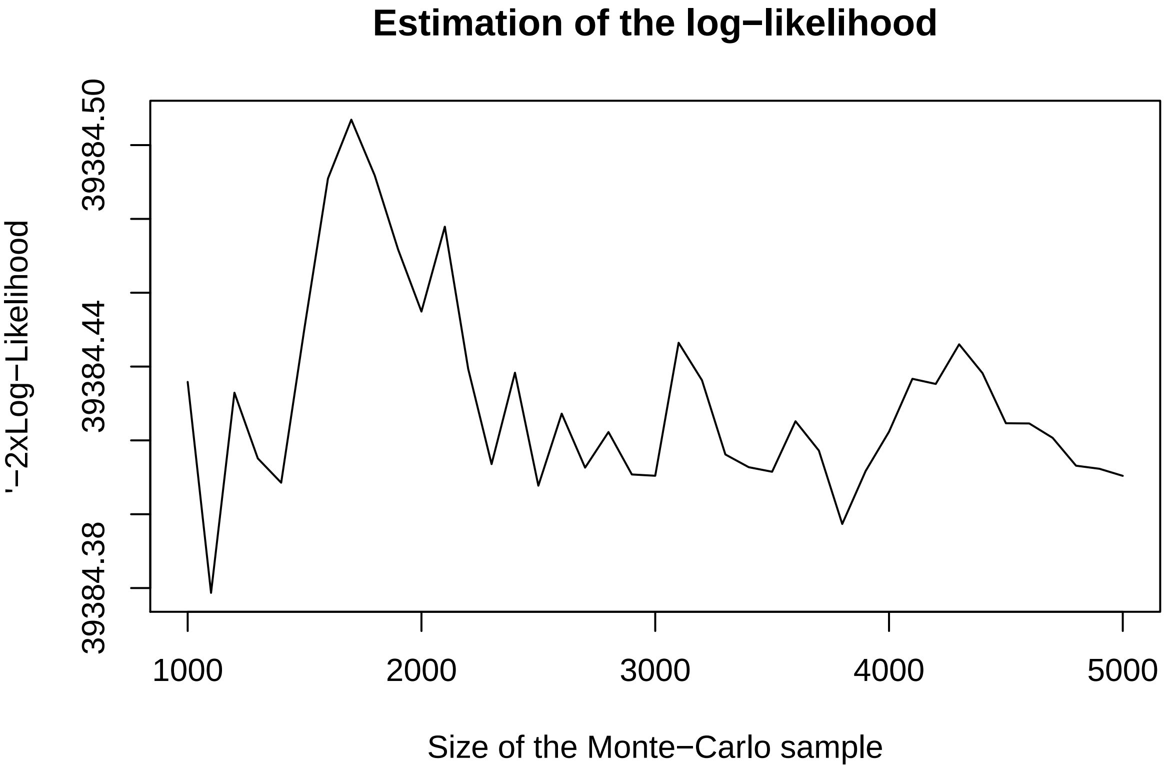


```
## --------------------------------------------------
## -----------------  Fixed effects  ------------------
## --------------------------------------------------
##      Parameter Estimate
## [1,] alp1      12.4
## [2,] alp2       1.4
## [3,] alp3       2.7
## [4,] beta      -5.7
## --------------------------------------------------
## -----------  Variance of random effects  -----------
## --------------------------------------------------
##      Parameter   Estimate
## alp1 omega2.alp1 37
## --------------------------------------------------
## ------  Correlation matrix of random effects  ------
## --------------------------------------------------
##              omega2.alp1
## omega2.alp1 1
## --------------------------------------------------
## ---------------  Statistical criteria  -------------
## --------------------------------------------------
##
## Likelihood computed by importance sampling
##       -2LL= 2565.985
##       AIC = 2577.985
##       BIC = 2607.431
## --------------------------------------------------
## Comparing estimates from saemix and Monolix

##                saemix       mlx     seMLX
## th1_pop   12.383990 12.013118 0.5371030
## th2_pop    1.355157  1.353616 0.1062970
## th3_pop    2.655177  2.594388 0.1840582
## th4_pop   -5.663764 -5.082603 0.2171057
## omega_th1  6.049610  6.204895 0.3890135
```

Count data

Note: dummy parameter only there to ensure we have at least 2 parameters to work on, but I thought we had fixed this…

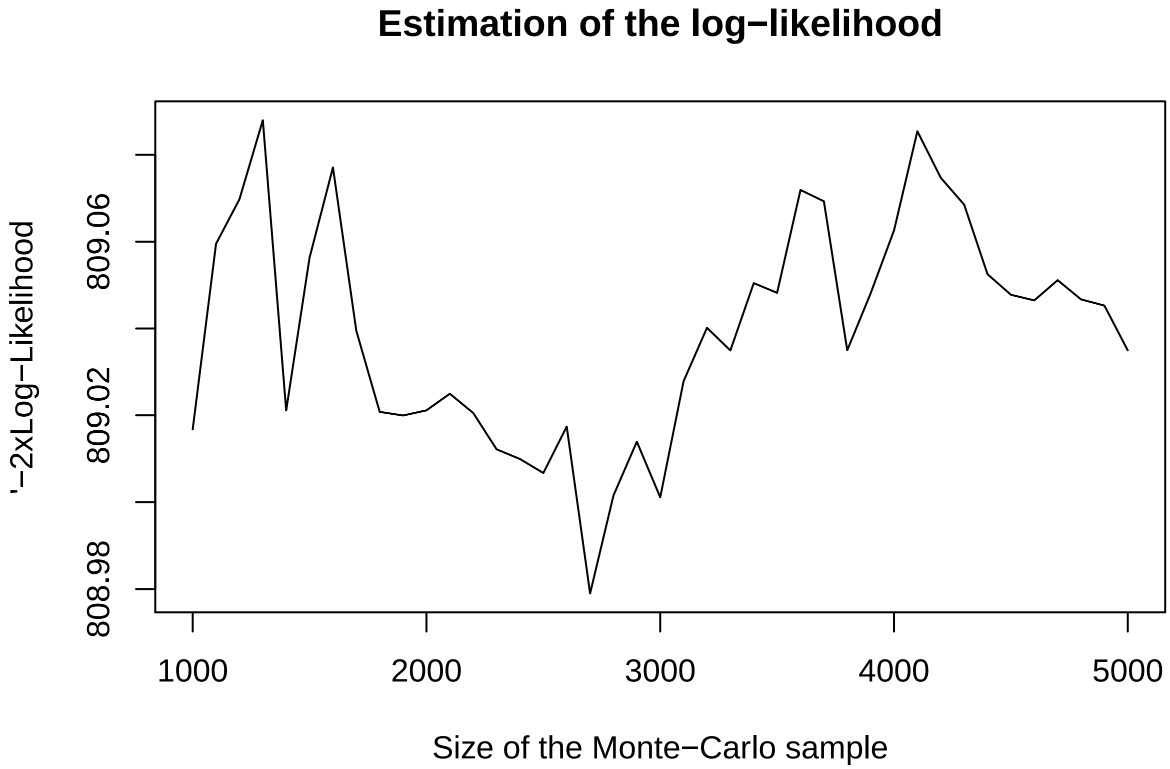


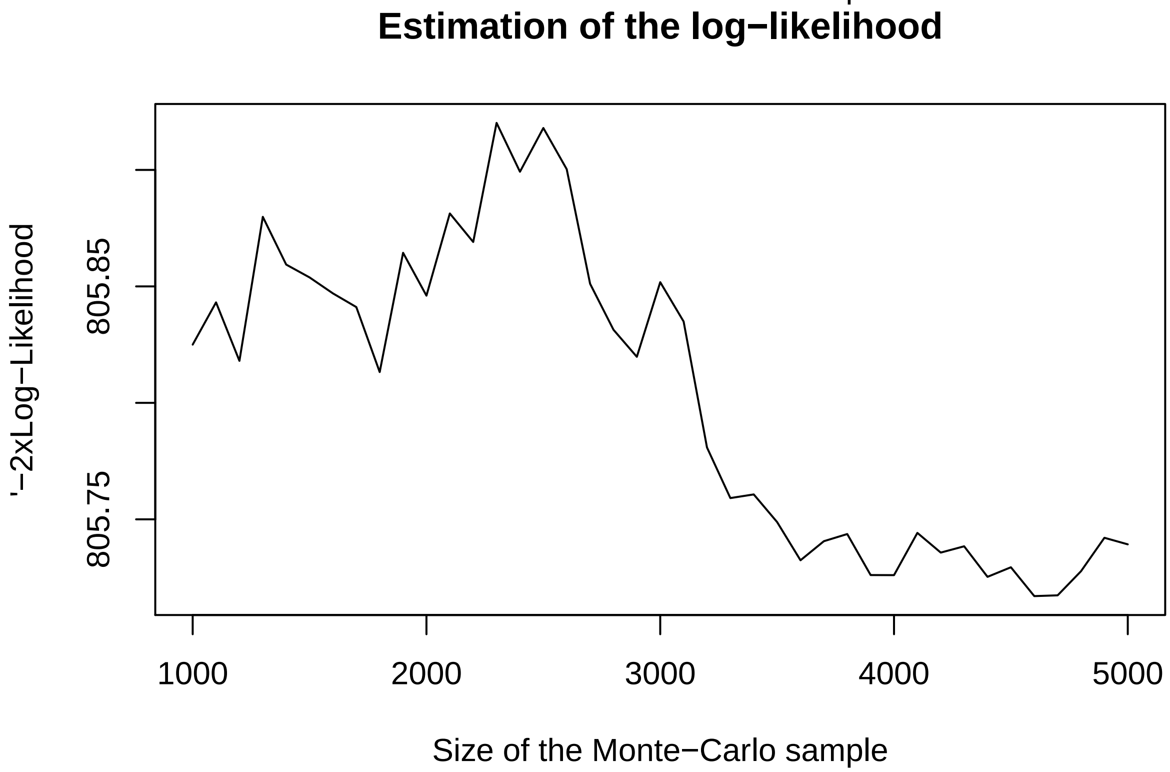
```
## ----------------------------------------------------
## -----------------  Fixed effects  ------------------
## ----------------------------------------------------
##      Parameter Estimate
## [1,] lambda    0.44
## [2,] dummy     1.00
## ----------------------------------------------------
## -----------  Variance of random effects  -----------
## ----------------------------------------------------
##        Parameter     Estimate
## lambda omega2.lambda 0.92
## ----------------------------------------------------
## ------  Correlation matrix of random effects  ------
## ----------------------------------------------------
##               omega2.lambda
## omega2.lambda 1
## ----------------------------------------------------
## ---------------  Statistical criteria  -------------
## ----------------------------------------------------
##
## Likelihood computed by importance sampling
##       -2LL= 39384.41
##       AIC = 39390.41
##       BIC = 39398.23
## ----------------------------------------------------
```

```
## Comparing estimates from saemix and Monolix
```

```
##                 saemix       mlx      seMLX
## lambda_pop   0.4447922 0.4721807 0.05282652
## omega_lambda 0.9566419 1.0614847 0.08711516
```

TTE data





```
## ---------------------------------------------------
## ----------------  Fixed effects  -----------------
## ---------------------------------------------------
##      Parameter Estimate
## [1,] lambda    42
## [2,] dummy      1
## ----------------------------------------------------
## -----------  Variance of random effects  -----------
## ----------------------------------------------------
##        Parameter     Estimate
## lambda omega2.lambda 0.4
## ----------------------------------------------------
## ------  Correlation matrix of random effects  ------
## ----------------------------------------------------
##               omega2.lambda
## omega2.lambda 1
## ----------------------------------------------------
## ---------------  Statistical criteria  -------------
## ----------------------------------------------------
##
## Likelihood computed by importance sampling
##       -2LL= 809.0349
##       AIC = 815.0349
##       BIC = 822.8505
## ----------------------------------------------------
```

```
## ----------------------------------------------------
## -----------------  Fixed effects  ------------------
## ----------------------------------------------------
##      Parameter Estimate
## [1,] lambda    38.9
## [2,] beta       1.4
## ----------------------------------------------------
## -----------  Variance of random effects  -----------
## ----------------------------------------------------
##        Parameter     Estimate
## lambda omega2.lambda 0.52
## ----------------------------------------------------
## ------  Correlation matrix of random effects  ------
## ----------------------------------------------------
##               omega2.lambda
## omega2.lambda 1
## ----------------------------------------------------
## ---------------  Statistical criteria  -------------
## ----------------------------------------------------
##
## Likelihood computed by importance sampling
##       -2LL= 805.7393
##       AIC = 813.7393
##       BIC = 824.16
## ----------------------------------------------------
```

```
## Comparing estimates from saemix and Monolix
```

```
##              saemix        mlx     seMLX
## Te_pop   42.2534112 42.0911165 4.4827138
## omega_Te  0.6296271  0.2420044 0.1598935
```

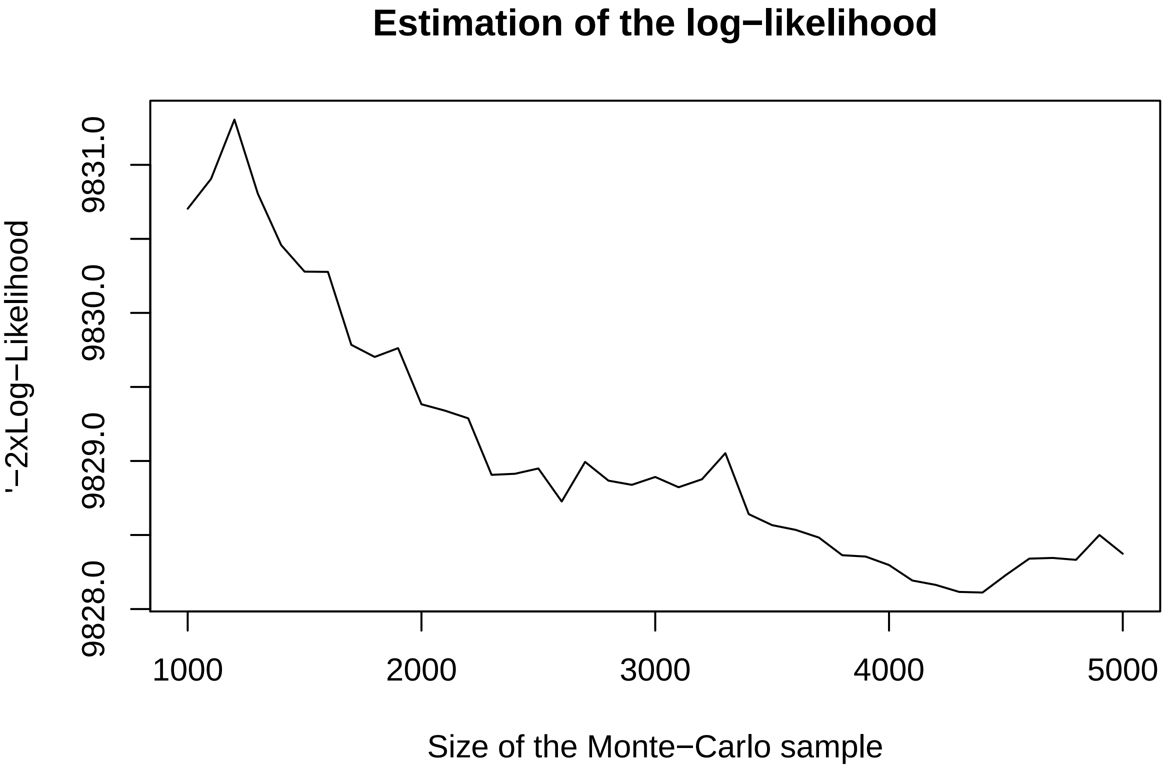
TODO: also compare the results for the Weibull example; here omega(Te) considerably higher with saemix compared to MLX, why ? (large SE on omegaTe according to monolix so maybe not very well estimated, but in another example I also saw something similar, saemix had a larger omega when that parameter had low information measured by large SE)

RTTE data

TODO

5. Example Ana

- saemix
  - works when using normal distributions for all parameters by mistake; parameter estimates *very* different from the simulated parameters
  - when using log-normal distributions for delta and gamma, problem to initialise
    - computing delta using the observed number of events in the dataset always gives (small) negative values (around -0.02 to -0.07)
    - computing gamma using the observed number of events in the dataset
- MLX
  - exactly same results when running through lixoftConnectors or using the Monolix interface
  - with the same mistake as above, gives different parameters than saemix but in the same area, and also quite different from the population parameters simulated
  - with the same structure as in the simulation (c(beta0_pop=1, o_beta0=0.3, gamma0_pop= 0.5, o_gamma0=0.3, delta0_pop=1, o_delta0=0.2), estimates are c(-0.305, 0.000879, 0.038) with omega=c(0.13,0.22,0.46), completely different from simulated parameters



Comparing estimates for Monolix and for saemix shows again very similar estimates:

```
## ----------------------------------------------------
## -----------------  Fixed effects  ------------------
## ----------------------------------------------------
##      Parameter Estimate
## [1,] beta      -0.26184
## [2,] gamma      0.00087
## [3,] delta      0.03421
## ----------------------------------------------------
## -----------  Variance of random effects  -----------
## ----------------------------------------------------
##       Parameter    Estimate
## beta  omega2.beta  0.037
## gamma omega2.gamma 0.073
## delta omega2.delta 0.364
## ----------------------------------------------------
## ------  Correlation matrix of random effects  ------
## ----------------------------------------------------
##              omega2.beta omega2.gamma omega2.delta
## omega2.beta  1           0            0
## omega2.gamma 0           1            0
## omega2.delta 0           0            1
## ----------------------------------------------------
## ---------------  Statistical criteria  -------------
## ----------------------------------------------------
##
## Likelihood computed by importance sampling
##       -2LL= 9828.373
##       AIC = 9842.373
##       BIC = 9879.566
## ----------------------------------------------------
```

```
## Comparing estimates from saemix and Monolix
```

```
##                     saemix          mlx        seMLX
## beta0_pop    -0.2618361255 -0.305302302 2.020247e-02
## gamma0_pop    0.0008728841  0.000879058 3.596399e-05
## delta0_pop    0.0342088071  0.037961374 1.655779e-03
## omega_beta0   0.1910939073  0.131080430 1.614692e-02
## omega_gamma0  0.2704689169  0.219729943 4.453560e-02
## omega_delta0  0.6036676380  0.463970748 5.891716e-02
```

The problem is therefore likely to lie in the design (informativeness?) or the simulation itself. When we use an R function to compute the probabilities with the parameters given in the simulation, the predicted probabilities turn out to be :

- nearly 1 (0.98) for the first simulated time (time=-10, when all doses are 0) => nearly all simulated events are Y=1
- nearly 0 (10^{-6} to 10^{-11} for any other time) => (nearly) all simulated events are Y=0

With such a design it would be impossible to estimate the parameters of the model anyway (impossible to separate time and dose effects).

```
## nobs
##    5
## 1500
```

```
##    dose times       P(Y=1)
## 13    0   -10 9.820138e-01
## 1    10     5 1.370957e-06
## 2    20     5 6.224145e-11
## 3    30     5 2.825757e-15
## 4    10    20 7.582560e-10
## 5    20    20 3.442477e-14
## 6    30    20 1.562882e-18
## 7    10    35 4.193796e-13
## 8    20    35 1.903980e-17
## 9    30    35 8.644057e-22
## 10   10    50 2.319523e-16
## 11   20    50 1.053062e-20
## 12   30    50 4.780893e-25
```

```
##
## 1 new("nonstandardGenericFunction", .Data = function (object, type = c("mode",
## 2     "mean"))
## 3 {
## 4     standardGeneric("psi")
## 5 }, generic = structure("psi", package = ".GlobalEnv"), package = ".GlobalEnv",
## 6     group = list(), valueClass = character(0), signature = c("object",
```

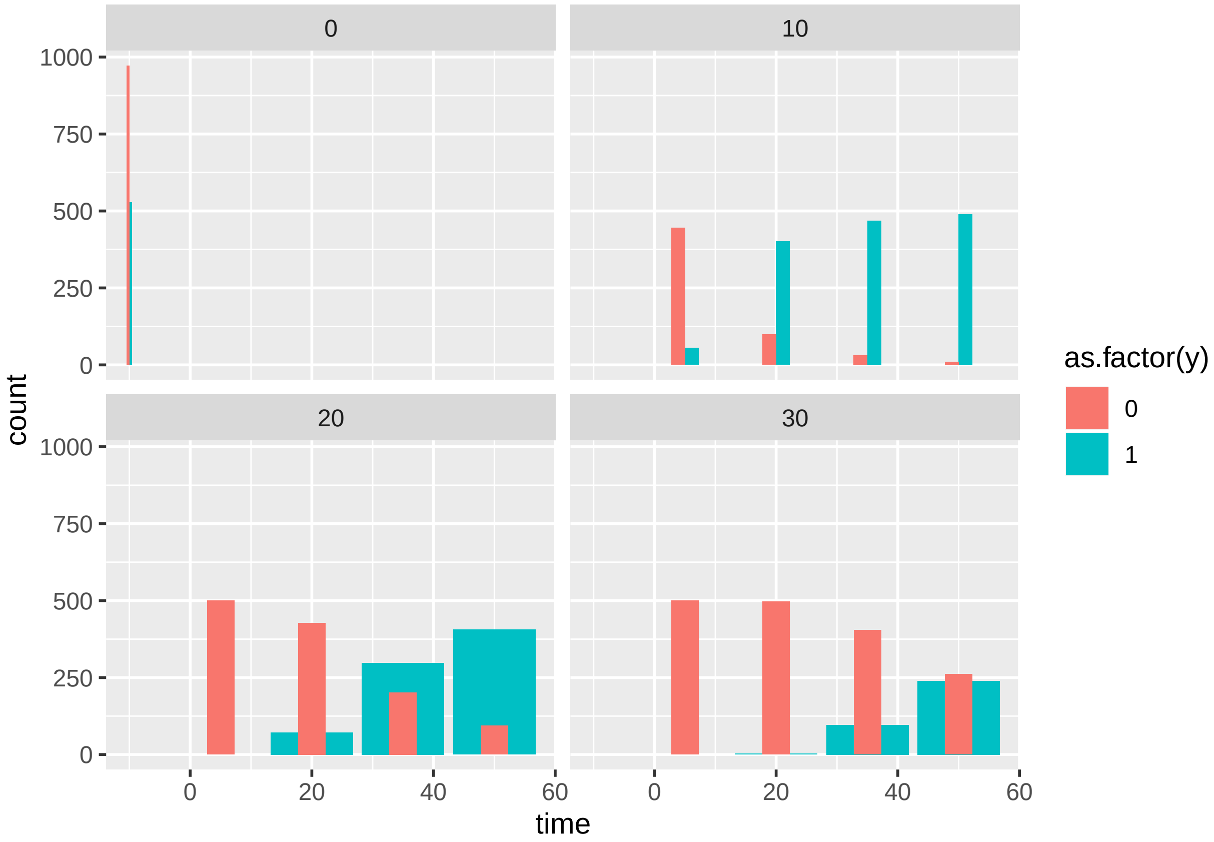
```
## Using individual parameters
```

```
## [1] 1437
```

```
## Frequency of simulated events at t=(-10): 0.958
```

```
## Frequency of simulated events at t>0: 0
```

Oddly enough though, this is not the data we have been given, as the data plot below shows. In fact, we clearly see the proportion of events at time=-10 is in the order of 0.3, we also see an increasing proportion of events as time increases, and a negative dose effect (less events for larger doses). This does not seem to be in line either with the simulated values, with both γ and δ given positive values (and log-normal distributions), whereas here we clearly see the time effect β should be negative (since it increases logit(P(Y=0)) and therefore decreases logit(P(Y=1))).



Conclusion: something is wrong either in the dataset I was given (does not correspond to the simulated file) or in simulx for this model => next step: run the simulx simulation and compare the output to the original dataset.

2. Tests from testbelhal

TTE data

3. Documentation examples

4. Marilou SIR

For Marilou 3 settings investigated (50, 100 and 224 subjects, with 2 treatment groups), using the N=100 setting as the results were starting to be correct at this stage.

5. Debug Ana