Joint Modelling of Longitudinal Mixed Effects and Accelerated Failure Time Model

Koushik Khan

M.Sc.-Sem-IV, Roll No.: MSC (Sem-IV) - Stat - 223

Department of Statistics

Visva-Bharati

Santiniketan, India

Outline

- Some prerequisites
- What is ***Joint Model (JM)***?
- Models for the individual components
- Construction of the likelihood
- Monte Carlo EM (MCEM) based estimation technique
- MCEM algorithm in a nutshell
- Simulation and analysis of a real data set
- Discussions
- References

From now I will call it JM

Some Prerequisites

- **What is longitudinal data ?**

Repeated measurements of one or more variable, taken on several (small in number) time points for many subjects, involved in an experiment, constitute a longitudinal data.

- **Example:**

i. Weekly lowest temperature of 50 largest cities in a month.
ii. Daily blood pressure measurements of 25 patients (say) during a week.

- ***Note: Since measurements are taken on the same subject (and each subject thereafter, such observations are correlated)***

- **What is time-to-event ?**

As the name suggests, it is a variable denoting the time needed up to the occurrence of an event.

- **Example:**

i. Time needed to fail for an electronic device.
ii. Time to recover from a disease.
iii. Duration of life before death of a patient after receiving a drug.

- ***Note: Time-to-event r.v. should have distribution with positive range***

Definition & classification of JM

- Suppose we observe two continuous processes e.g. longitudinal Y and time-to-event T. By joint modelling we mean to consider their joint likelihood based on available data, to find the estimates of the parameters involved.
- There are two different strategies to factorize the joint density of (Y, T) (Little (1995)) based on model interpretations, and consequently, suitability for individual problems.
- These are ***Selection Model*** and ***Pattern Mixture Model***

$$\textit{Selection model} \qquad \textit{Pattern mixture model}$$
$$[\boldsymbol{Y},\boldsymbol{T}] = [\boldsymbol{T}|\boldsymbol{Y}][\boldsymbol{Y}] \qquad [\boldsymbol{Y},\boldsymbol{T}] = [\boldsymbol{Y}|\boldsymbol{T}][\boldsymbol{T}]$$

- Selection model would answer the question regarding how one's response on the severity of the disease affects death (failure).
- ***Selection model*** will be used for rest of the discussion.

Justification & Area of application

- In clinical & other follow-up studies longitudinal data and survival data frequently arise together.
- For example, collecting information on blood pressure repeatedly over time and recording the time to recover form a disease for several patients.
- Logical to think that these two processes are associated in some ways.
- Separate analyses may lead to inefficient inferences.
- Joint models of longitudinal and survival data, on the other hand, incorporate all information simultaneously and may provide valid and efficient inferences.

Specifications of the Models

❑ Longitudinal model:

❑ It is often modelled as,

Liner mixed-effects model

$$y_{ij} = \mu + \alpha x_{ij} + b_i + \epsilon_{ij}; \quad 1 \leq j \leq n_i, \quad 1 \leq i \leq N$$

N: Number of subjects involved
n_i: Number of longitudinal observation
x_{ij}: Value of a single covariate
b_i: Shared random effect, $\sim N(0, \sigma_b^2)$
ϵ_{ij}: Standardised random error, $\sim N(0, 1)$

AFT model with log-normal distn.

❑ Time-to-event model:

$$\log T_{ij} = \beta + b_i$$

T_{ij}: $min\left(T^{\star}_{ij}, C_{ij}\right)$, $T^{\star}_{ij}$: Actual time
b_i: Shared random effect, C_{ij}: Censoring cut-off

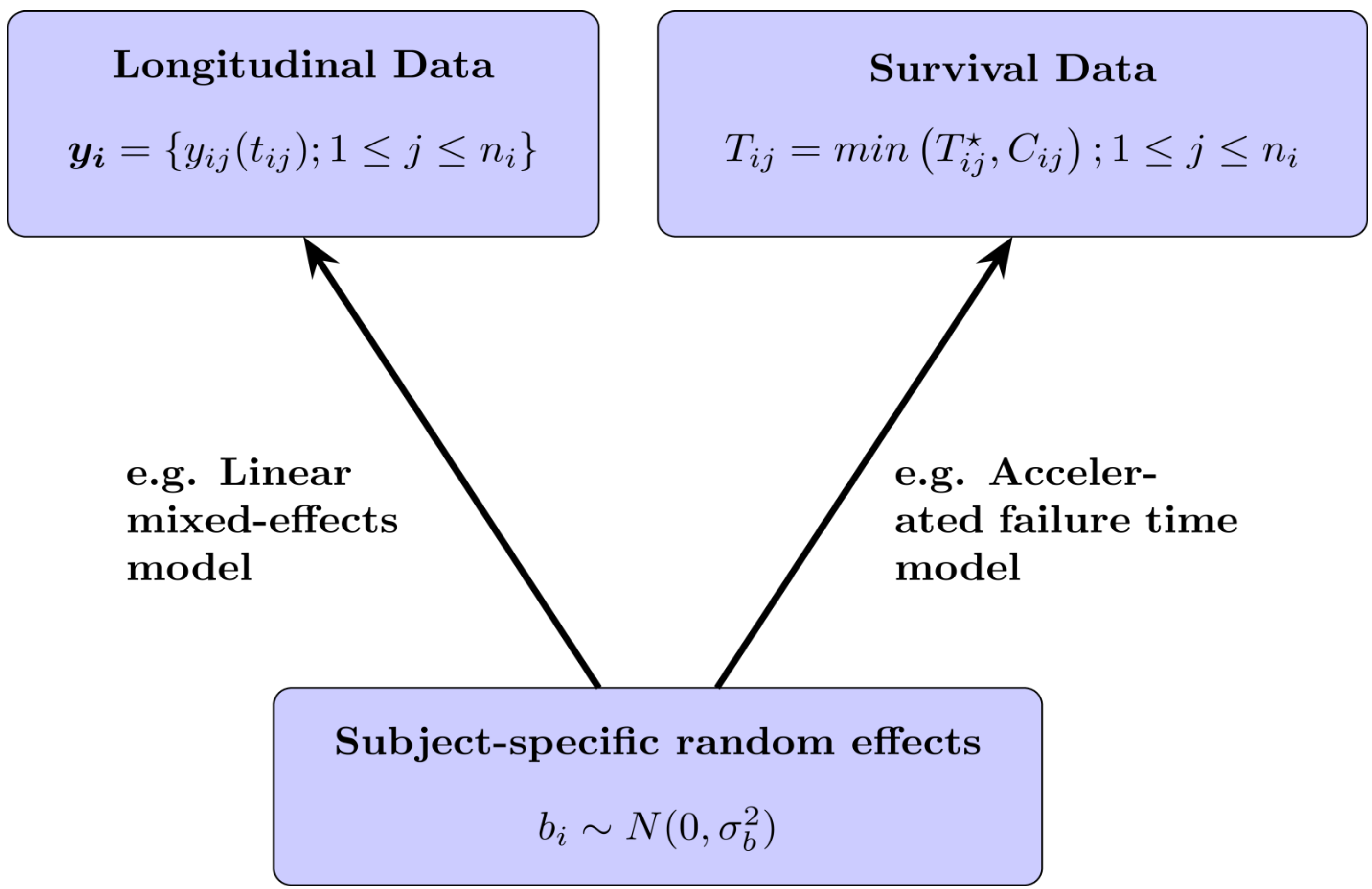
$N((\beta + b_i), \sigma_T^2)$

we define,

$$\begin{aligned} \delta_{ij} &= 1; \quad \text{if actual event time is recorded} \\ &= 0; \quad \text{if event time is censored to the right} \end{aligned}$$

Roll of random effects

- Longitudinal observations, for a particular subject, are necessarily dependent.
- Random effects are incorporated to make the longitudinal outcomes as well as the two models dependent.



Construction of the joint likelihood

- Joint likelihood of longitudinal and time-to-event for ith subject is given by,

$$
\begin{aligned}
L_i(\boldsymbol{\theta}) &= \prod_{j=1}^{n_i} f(y_{ij}, \log t_{ij} | \boldsymbol{\theta}) \\
&= \int_{b_i} \prod_{j=1}^{n_i} g_1(\log t_{ij} | y_{ij}, b_i) \times g_2(y_{ij} | b_i) \times \pi(b_i) db_i \\
&= \int_{b_i} \prod_{j=1}^{n_i} g_1(\log t_{ij} | b_i) \times g_2(y_{ij} | b_i) \times \pi(b_i) db_i \\
&= \int_{b_i} l_{iJ} db_i
\end{aligned}
$$

- Finally,

$$
L(\boldsymbol{\theta}) = \prod_{i=1}^{N} L_i(\boldsymbol{\theta})
$$

Continued...

- It can be shown that,

$$l_{iJ} \propto \prod_{j=1}^{n_i} \left\{ \frac{1}{\sigma_T} \phi \left(\frac{\log t_{ij} - \beta - b_i}{\sigma_T} \right) \right\}^{\delta_{ij}} \times \left\{ 1 - \Phi \left(\frac{\log c_{ij} - \beta - b_i}{\sigma_T} \right) \right\}^{1-\delta_{ij}} \times \phi \left(y_{ij} - \mu - \alpha x_{ij} - b_i \right) \times \frac{1}{\sigma_b} \phi \left(\frac{b_i}{\sigma_b} \right)$$

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \int_{b_i} l_{iJ} db_i$$

too complicated to handle mathematically

but actually we need this

Needs numerical optimization techniques !

MCEM based estimation technique, a tricky way out !

- We use *Fisher's Scoring* method to get the m.l.e of $\boldsymbol{\theta} = (\mu, \alpha, \beta, \sigma_b)'$
- **Score vector:**

$$\begin{aligned} \boldsymbol{S_\theta} &= \frac{\partial}{\partial \boldsymbol{\theta}} \log L(\boldsymbol{\theta}) \\ &= \sum_{i=1}^{N} \left\{ \int_{b_i} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log l_{iJ} \right) \times \frac{l_{iJ}}{\int_{b_i} l_{iJ} db_i} db_i \right\} \\ &= \sum_{i=1}^{N} \left\{ \int_{b_i} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log l_{iJ} \right) \times h\left(b_i | y_{ij}, \log t_{ij}\right) db_i \right\} \end{aligned}$$

condn. density of $\boldsymbol{b_i | y_{ij}, \log t_{ij}}$

So that,

$$\begin{aligned} \boldsymbol{S_\theta} &= \sum_{i=1}^{N} E_{b_i | y_{ij}, \log t_{ij}} \left(\frac{\partial}{\partial \boldsymbol{\theta}} \log l_{iJ} \right) \\ &= \sum_{i=1}^{N} E_{b_i | \text{data}} \left(\boldsymbol{S^{\star}_{\theta i}} \right) \end{aligned}$$

condn. score vector

- **Information matrix:**

$$\boldsymbol{I}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \boldsymbol{I}(\boldsymbol{\theta})^{\star}_{\boldsymbol{i}}$$

condn. information matrix

But how to compute the expectations ?

- It can't be always computed exactly !
- But can be approximated.

Thanks to Markov Chains Monte Carlo Technique !

- First we simulate random observations from $h\left(b_i|y_{ij},\log t_{ij}\right)$ using **Metropolis Sampler**
- Use these observations to approximate the expectation as follows:

$$E_{b_i|y_{ij},\log t_{ij}}\left(\boldsymbol{S}^{\star}_{\boldsymbol{\theta i}}\right) \simeq \frac{1}{(R-burnin)} \sum_{\ell=1}^{(R-burnin)} \boldsymbol{S}^{\star}_{\boldsymbol{\theta i}}{}^{(\ell)}$$

- Hence,

$$\boldsymbol{S}_{\boldsymbol{\theta}} \simeq \sum_{i=1}^{N} \frac{1}{(R-burnin)} \sum_{\ell=1}^{(R-burnin)} \boldsymbol{S}^{\star}_{\boldsymbol{\theta i}}{}^{(\ell)}$$

Iterative Maximization

- Use iterative maximization algorithm successively to maximize that **complicated likelihood**

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \boldsymbol{I}\left(\boldsymbol{\theta}^{(\boldsymbol{k})}\right)^{\boldsymbol{-1}} \boldsymbol{S}\left(\boldsymbol{\theta}^{(\boldsymbol{k})}\right)$$

Fisher's Scoring Method

- **And wait for convergence**

MCEM algorithm in a nutshell

Step 1. Choose an initial estimate of θ, say, $\theta^{(0)}$

(using some suitable techniques)

Step 2. Set k = 0

Step 3. Generate R (10000, say) observations from $h(b_i|y_{ij}, log(t_{ij}))$ using Metropolis Sampler (a lot of difficulties involved!) taking some suitable proposal (*Markov Chain*)

Step 4. Approximate the expectations involved in score function and information matrix (*E-step withMonte Carlo*)

Step 5. Update the estimate as,

$$\theta^{(k+1)} = \theta^{(k)} + \left[I\left(\theta^{(k)}\right)\right]^{-1} S\left(\theta^{(k)}\right)$$

(*M-step with Fisher's scoring*)

Step 6. Set k = k + 1

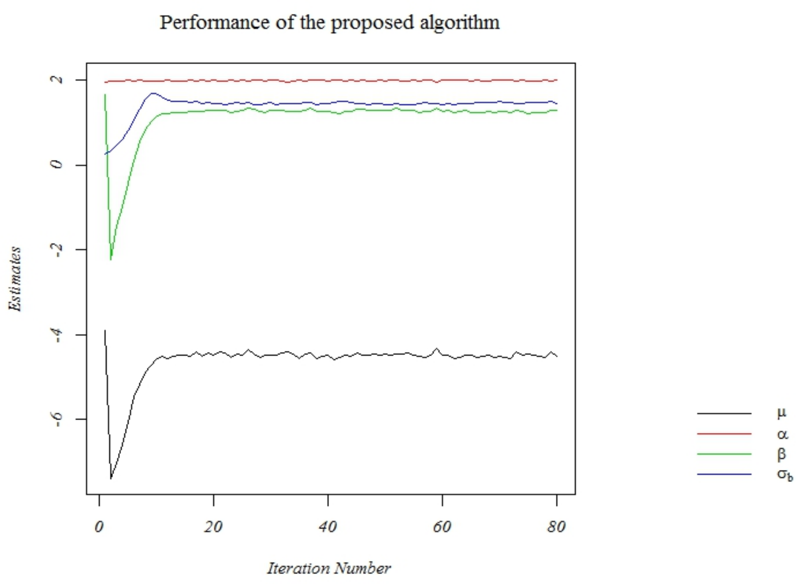
Step 7. Continue **steps 2-6** till the convergence

Simulation Studies

- Simulation studies are performed for three choices of σ_T. N & n are kept fixed at 25 & 4 respectively.
- **Study 1: Data generated by taking $\sigma_T = 1.55$**

Table 1: Parameter estimates obtained using MCEM algorithm

Parameter	True value	Estimate	Standard Error	Absolute bias
μ	-3	-4.504973	0.53672034	1.504973
α	2	2.010104	0.01294196	0.010104
β	2.5	1.297291	0.59741468	1.202709
σ_b	1.225	1.467351	0.25016115	0.242606



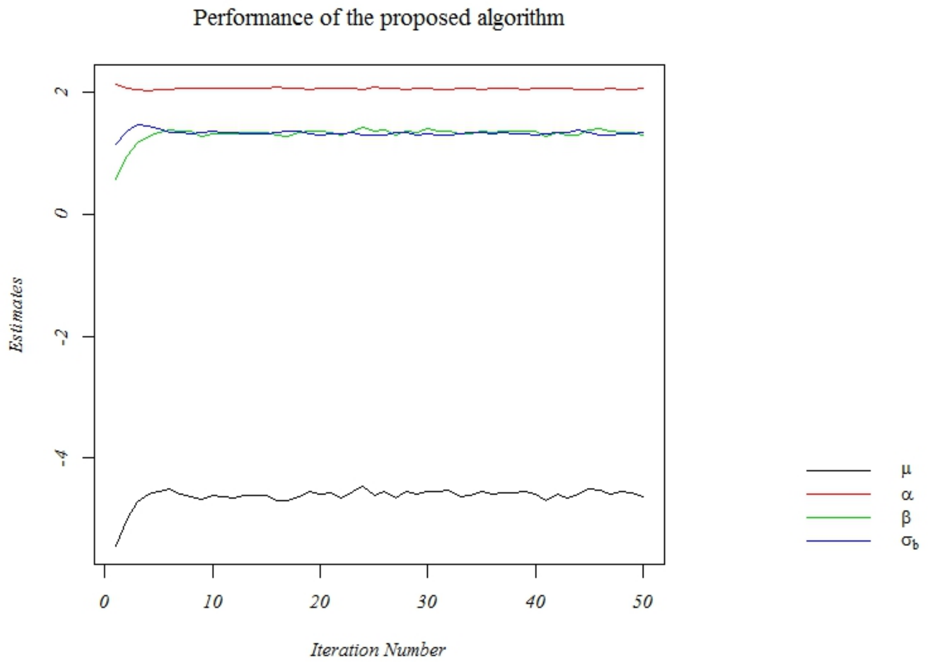
- ✓ # of iteration: 80
- ✓ Estimate of α shows almost constant pattern.
- ✓ α&σ_b has biases closed to zero.

Continued…

- **Study 2: Data generated by taking $\sigma_T = 0.50$**

Table 1: Parameter estimates obtained using MCEM algorithm

Parameter	True value	Estimate	Standard Error	Absolute bias
μ	-3	-4.635835	0.14281614	1.6358
α	2	2.061284	0.01618356	0.061284
β	2.5	1.294264	0.12623253	1.2057
σ_b	1.225	1.346767	0.04300177	0.1218



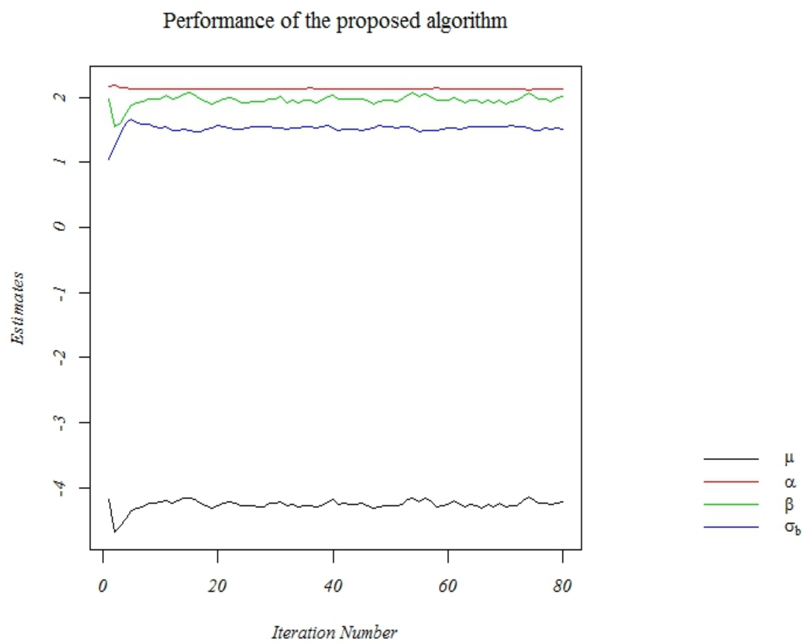
- ✓ # of iteration: 50
- ✓ Estimate of α shows usual pattern .
- ✓ now μ, β *and* σ_b has less fluctuations near the starting area

Continued...

- **Study 3: Data generated by taking $\sigma_T = 0.15$**

Table 1: Parameter estimates obtained using MCEM algorithm

Parameter	True value	Estimate	Standard Error	Absolute bias
μ	-3	-4.229542	0.07972108	1.2295
α	2	2.132087	0.01024069	0.132087
β	2.5	2.002448	0.07776745	0.4976
σ_b	1.225	1.507479	0.07079735	0.2857



- ✓ # of iteration: 80
- ✓ μ behaves as usual
- ✓ after 7 or 8 iterations σ_b is more or less constant

Analysis of Muscular Dystrophy Syndrome data

- Data set consists of observations on composite muscle score (i.e. average of 10 muscle scores, which are responsible for walking).
- Two time to event indicators are also observed over different time points- time taken to walk 4 steps and time to get up from lying state.
- Objective is to characterize the relationship between failure time and the longitudinal outcomes.
- Only **time to walk 4 steps** is considered (advised by the doctors) in the present work.

- Data Description click here

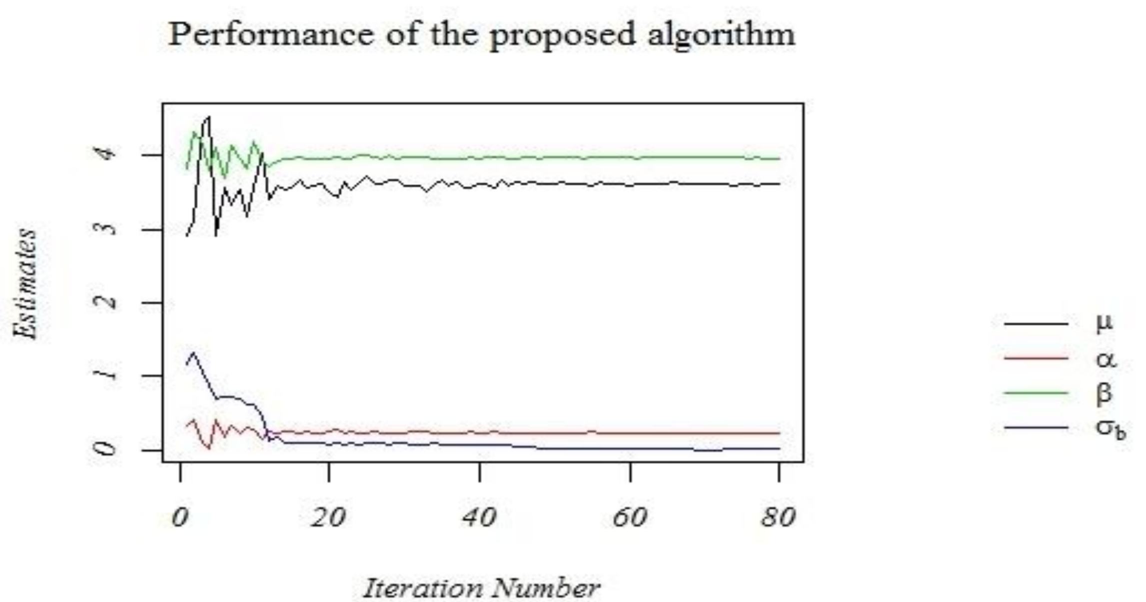
Continued…

- Estimates of the parameters with standard errors are given below:

Table 1: Parameter estimates applying the Joint Model

Parameter	Estimate	Standard Error
μ	3.61647686	0.20449250
α	0.23480275	0.04372608
β	3.96921363	0.07389875
σ_b	0.01304504	0.28348920

- Performance of the MCEM algorithm for real data is shown in the following figure:



Discussion

- Principal objective of this work is to obtain the estimates of the parameters for a data set on **muscular dystrophy syndrome**.
- We have noticed that MCEM performed well for fixed values of σ_T i.e. the standard deviation of time-to-event variable.
- But when considering σ_T as random, sometimes the algorithm produces negative estimate of σ_T .
- Random errors in longitudinal model are standardized.
- In longitudinal model we have incorporated only one covariate, however more than covariate may also be used.
- Instead of linear, some suitable non-linear functions may also be used in both the models.

Some Selected References

- Chakraborty, A. and Das, K. (2010). Inferences for joint modelling of repeated ordinal scores and time to event data. Computational and Mathematical Methods in Medicine, **11**, 281-295
- Little, R. (1995). Modeling the dropout mechanism in repeated measures studies. Journal of the American Statistical Association, **90**, 1112-1121.
- Guo, X. and Carlin, B.P. (2004). Separate and joint modelling of longitudinal and event time data using standard computer packages. The American Statistician, **58**, 16-24.
- Henderson, R., Diggle, P.J. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data, Biometrics, **1**, 465-80.
- Hogan, J. and Laird, N. (1997). Mixture models for the joint distribution of repeated measures and event times. Statistics in Medicine, **16**, 239-258.
- Pawitan, Y. and Self, S. (1993). Modelling disease marker processes in aids. Journal of the American Statistical Association, **88**, 719-726.
- Schluchter, M.D. (1992). Methods for the analysis of informatively censored longitudinal dada. Statistics in Medicine, **11**, 1861-1870.

Thank You !