Score test for random changepoint in a mixed model

Corentin Segalas and Hélène Jacqmin-Gadda
INSERM U1219, Biostatistics team, Bordeaux

GDR Statistiques et Santé

October 6, 2017

BORDEAUX POPULATION HEALTH Research Center - U1219 | Biostatistics

Alzheimer's Disease (AD)

- A major public health issue today and tomorrow
- A very long pre-diagnostic phase
- Heterogeneous and non-linear decline trajectories

Different profiles?

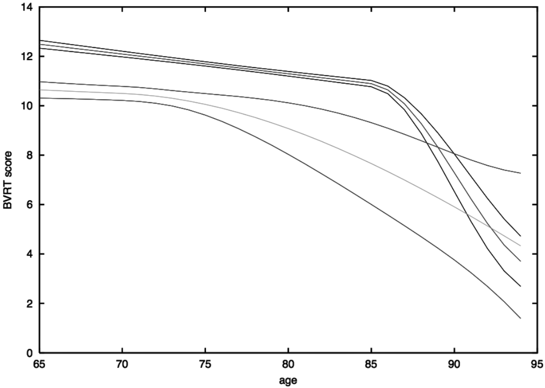


Figure: Estimated mean BVRT score according to age for 2 subjects demented at 90 with low or high educational level (Jacqmin-Gadda et al., 2006)

Objective

Propose a test for the existence of a random changepoint in a mixed model for longitudinal data.

The mixed model with random changepoint

$$Y(t_{ij}) = Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_2 \sqrt{(t_{ij} - \tau_i)^2 + \gamma} + \varepsilon_{ij} \qquad (1)$$

with

$\beta_{ki} = \beta_k^T X_{ki} + \alpha_{ki}$ for $k = 0, 1$,

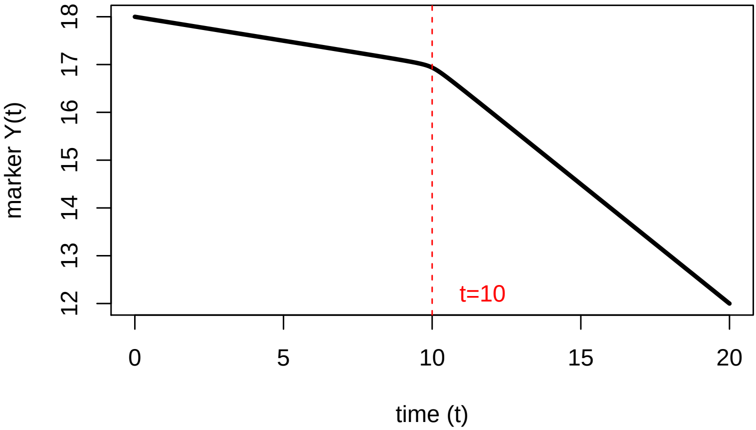
$\alpha_i = (\alpha_{0i}, \alpha_{1i})^T \sim \mathcal{N}(0, B)$,

$\tau_i = \mu_\tau + \sigma_\tau \tilde{\tau}_i$ with α_i independent from $\tilde{\tau}_i$ and $\tilde{\tau}_i \sim \mathcal{N}(0, 1)$,

$\gamma = 0.1$ a fixed smoothness parameter.

β_{1i} is the mean slope and β_2 half the difference of the slopes.

The mixed model with random changepoint



Estimation

- Model estimated by MLE and integral computed by gaussian quadrature (15 nodes)

$$\ell_n(Y;\theta) = \sum_{i=1}^{n} \log \iint \prod_{j=1}^{n_i} f(Y_{ij}|\alpha_i, \tilde{\tau}_i) f(\alpha_i) f(\tilde{\tau}_i) \mathrm{d}\alpha_i \mathrm{d}\tilde{\tau}_i.$$

- No known methods to test the existence of a *random* CP

Classic score test

- $(H_0) : \beta_2 = \beta_2^0$ vs. $(H_1) : \beta_2 \neq \beta_2^0$
- test statistic:

$$S_n = \frac{U_n(\beta_2^0, \hat{\theta}_0)^2}{Var(U_n(\beta_2^0, \hat{\theta}_0)^2)} \text{ with } U_n(\beta_2^0, \hat{\theta}_0) = \left. \frac{\partial \ell_n(Y; \beta_2, \hat{\theta}_0)}{\partial \beta_2} \right|_{\beta_2 = \beta_2^0}$$

  with $\hat{\theta}_0$ the MLE of nuisance parameters under the null
- null distribution: $\chi^2(1)$

Identifiability issue

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_2 \sqrt{(t_{ij} - \mu_\tau - \sigma_\tau \tilde{\tau}_i)^2 + \gamma} + \varepsilon_{ij}$$

Hypotheses:

$$(H_0) : \beta_2 = 0 \text{ vs. } (H_1) : \beta_2 \neq 0$$

- nuisance parameters : $\beta_0, \beta_1, \sigma, \sigma_0, \sigma_1, \sigma_{01}, \mu_\tau, \sigma_\tau$
- μ_τ and σ_τ unidentifiable under the null: we can't use the classic score test statistic S_n which depends on them.

The score under the null ($\beta_2 = 0$)

$$U_n(0;\theta) = \sum_{i=1}^{N} \left[\int f(\tilde{\tau}_i) \int f(\alpha_i) \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(Y_{ij} - \beta_{0i} - \beta_{1i}t_{ij})^2 \right\} \mathrm{d}\alpha_i \mathrm{d}\tilde{\tau}_i \right]^{-1}$$

$$\times \iint f(\alpha_i) f(\tilde{\tau}_i)(\sqrt{2\pi}\sigma)^{-n_i} \sum_{j=1}^{n_i} \left[\frac{1}{\sigma^2} \exp\left\{ -\frac{1}{2\sigma^2}(Y_{ij} - \beta_{0i} - \beta_{1i}t_{ij})^2 \right\} \right.$$

$$\left. (Y_{ij} - \beta_{0i} - \beta_{1i}t_{ij}) \times \sqrt{(t_{ij} - \mu_\tau - \sigma_\tau \tilde{\tau}_i)^2 + \gamma} \prod_{k \neq j} \exp\left\{ -\frac{1}{2\sigma^2}(Y_{ik} - \beta_{0i} - \beta_{1i}t_{ik})^2 \right\} \right] \mathrm{d}\alpha_i \mathrm{d}\tilde{\tau}_i$$

How to circumvent this problem?

Score test with identifiability issue

Classic problem when testing homogeneity on mixture models.
Two main approaches :

- replace μ_τ and σ_τ by the MLE under the alternative (Conniffe, 2001)
- consider the supremum in (μ_τ, σ_τ) of the score test statistic (Hansen, 1996)

The sup score test

- $(H_0) : \beta_2 = 0$ vs. $(H_1) : \beta_2 \neq 0$
- test statistic:

$$T_n = \sup_{(\mu_\tau, \sigma_\tau)} S_n(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0)$$

with

$$S_n(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0) = \frac{U_n(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0)^2}{Var(U_n(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0))}$$

with $\hat{\theta}_0$ the MLE of identifiable nuisance parameters under the null

- null distribution: approached by MC perturbation algorithm or multiplier bootstrap (van der Vaart and Wellner, 1996).

The complete procedure

1. estimation of the null model (linear mixed model) using `nlme`
2. computing the observed test statistic (optimization via quasi-Newton and integral via pseudo-adaptive gaussian quadrature)

$$T_n^{obs} = \sup_{(\mu_\tau, \sigma_\tau)} \frac{U_n(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0)^2}{\hat{Var}(U_n(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0))}$$

where $U_n(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0) = \sum_{i=1}^n u_i(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0)$ and the variance is estimated by

$$\sum_{i=1}^n u_i(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0)^2.$$

The complete procedure

4. perturbation algorithm: for $k = 1, \ldots, K = 500$
   - generate n r.v. $\xi_i^{(k)} \sim \mathcal{N}(0,1)$
   - compute

$$T_n^{(k)} = \sup_{(\mu_\tau, \sigma_\tau)} \frac{\left(\sum_{i=1}^n u_i(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0) \xi_i^{(k)}\right)^2}{\sum_{i=1}^n u_i(0; \mu_\tau, \sigma_\tau, \hat{\theta}_0)^2}$$

5. compute the empirical $p-$value

$$p_K = \frac{1}{K} \sum_{k=1}^{K} \mathbf{1}_{T_n^{(k)} > T_n^{(obs)}}$$

Tests for the variability of β_2

If we reject the null hypothesis, we can test if there is a

1. random effect for the difference of slope:

$$\beta_{2i} = \beta_2 + \alpha_{2i} \text{ with } \alpha_i = (\alpha_{0i}, \alpha_{1i}, \alpha_{2i}) \sim \mathcal{N}(0, B)$$

$\Rightarrow$ corrected LR test for variance component (Stram and Lee, 1994)

2. dependance on covariates:

$$\beta_{2i} = \beta_{20} + \beta_{21} X_{2i}$$

$\Rightarrow$ Wald test

Simulation scenarios

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_2 \sqrt{(t_{ij} - \tau_i)^2 + \gamma} + \varepsilon_{ij}$$

with

$\beta_{0i} = 20 + \alpha_{0i}$ and $\beta_{1i} = -0.3 + \alpha_{1i}$

$\alpha_i = (\alpha_{0i}, \alpha_{1i})^T \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}\right),$

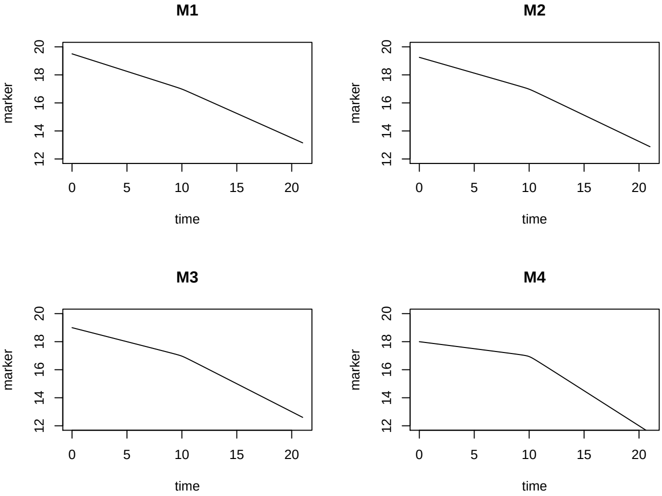
$\tau_i = 10 + 2\tilde{\tau}_i$ with α_i independant from $\tilde{\tau}_i$ and $\tilde{\tau}_i \sim \mathcal{N}(0,1)$,

$\gamma = 0.1, \sigma_\varepsilon = 1, t_{ij} = 0, 3, 6, 9, 12, 15, 18, 21$ for all i,

$\beta_2 = 0, -0.05, -0.075, -0.1, -0.2,$

Probability of drop-out at each visit: $0.1 \Rightarrow$ around 50% of the sample remaining at $t = 21$.

Simulation Scenarios



Results

N		100		200	
drop-out		no	yes	no	yes
size	M_0	0.029	0.042	0.034	0.039
power	M_1	0.397	0.054	-	-
	M_2	0.749	0.067	-	-
	M_3	0.949	0.093	-	-
	M_4	1	0.206	-	0.425

Table: Size and power of the test computed on 1000 replicates of each scenarios with $K = 500$ perturbations.

The PAQUID cohort

- 3777 subjects older than 65 from the french departments of Gironde and Dordogne, 25 years follow-up
- Marker : Isaac 15s score
- sample selection: incident case of dementia between year 1 and 25
- High education sample
  - 522 subjects with at least 1 measure
  - 1 to 12 measures by subject (mean = 5.8)
- Low education sample
  - 358 subjects with at least 1 measure
  - 1 to 12 measures by subject (mean = 4.6)
- model (1) with $\beta_{ki} = \beta_k + \alpha_{ki}$ for $k = 0, 1$ (no covariate)

Score test results

	obs. statistic test*	p-value
High education	14.059	0.001
Low education	1.388	0.443

Table: Score test results with $K = 1000$

For the high education subjects, we clearly reject the null hypothesis of no random changepoint.

Estimation (nq = 15)

	PAQUID demented sample			
	High education		Low education	
N	522		358	
Log-lik	-8845.889		-4685.284	
	Est	sd	Est	sd
β_0	23.087	0.219	20.477	0.342
β_1	-0.838	0.026	-0.531	0.039
β_2	-0.559	0.022	-0.354	0.033
μ_τ	-4.101	0.375	-5.512	0.694
σ	3.476	0.045	3.358	0.074
σ_0	4.195	0.134	3.873	0.172
σ_1	0.213	0.017	0.209	0.024
σ_τ	2.925	0.016	1.776	0.026
σ_{01}	0.275	0.343	0.178	0.675
slope 1/2	-0.279 / -1.397		-0.177 / -0.885	

Estimation of the mixed model with random CP

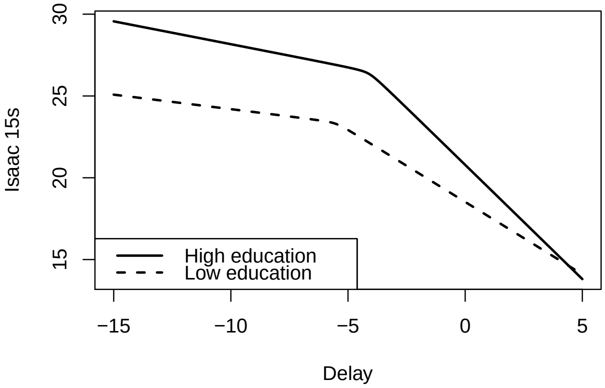


Figure: Mean estimation trajectory of the mixed model with random changepoint on the two educational level subsamples.

Variability of β_2: random effect ?

On high education subsample

$$(H_0) : \sigma_2 = 0 \text{ vs. } (H_1) : \sigma_2 \neq 0$$

where $\beta_{2i} = \beta_2 + \alpha_{2i}$ with $\alpha_{2i} \sim \mathcal{N}(0, \sigma_2^2)$.
LRS = -137.2 $\Rightarrow p < 0.001$
$\Rightarrow$ We need to add a random effect on β_2

Next steps

- simulations with varying σ_τ
- extension to :
  - joint models
  - joint multi-state models for interval censored data
  - models for multiple markers
  - etc.

References

1. Conniffe, D., "Score Tests When a Nuisance Parameter Is Unidentified under the Null Hypothesis." Journal of Statistical Planning and Inference (2001)
2. Hansen, Bruce E., "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesi." Econometrica (1996)
3. Jacqmin-Gadda, H., Commenges, D. and Dartigues, J.-F., "Random changepoint model for joint modeling of cognitive decline and dementia." Biometrics (2006)
4. Stram, D. O., and Lee J.W., "Variance Components Testing in the Longitudinal Mixed Effects Model." Biometrics (1994)
5. van der Vaart, A. W. and Wellner, J.A., "Weak Convergence and Empirical Processes." Chapter 2.9, Springer Series in Statistics (1996).

Thank you for
your attention!