Software for the joint modelling of longitudinal and survival data: the JoineR package

Pete Philipson

Collaborative work with Ruwanthi Kolamunnage-Dona, Inês Sousa, Peter Diggle, Rob Henderson, Paula Williamson & Gerwyn Green

useR! conference 2010, NIST, Gaithersburg, MD

Outline

- Longitudinal and survival data
- Joint modelling
- The JoineR package
- Simulations and performance
- Application to real data: liver cirrhosis and CD4 cell counts
- Future work and plans

Longitudinal and survival data

- Longitudinal data
  - Focus on linear mixed-effects model

Longitudinal sub-model

$$Y_{ij} = X_{1i}\beta_1 + R_{1i}(t_{ij}) + \epsilon_{ij}$$

  - $R_1 = D_1 U_1$ with U_1 multivariate Gaussian random effects and D_1 a random effects design marix

- Survival data
  - Consider two alternatives for the event times F
    1. Cox proportional hazards

$$h_i(t) = h_0(t)\exp(X_{2i}\beta_2 + R_{2i})$$

    2. Transformed Gaussian

$$F \sim LN(\mu_F, \sigma_F^2)$$

Joint modelling

- Suitable for a range of objectives
  1. Analysing repeated measures Y in the presence of informative drop-out times F
  2. Analysis of survival times F acknowledging the association with Y, which may be a time-varying explanatory covariate subject to measurement error
  3. Relationship between Y and F is of joint interest
- Examples of two of these will be demonstrated later

Joint models

- Random effects (RE) joint model
  - Sub-models linked through common random effects U
  - Strength of association measured through parameter(s) γ, i.e. $R_2 = \gamma R_1$
  - Model fitting achieved via EM algorithm
- Transformation model
  - Sub-models formulated as multivariate Gaussian

$$(Y, \log F) \sim MVN(\mu, \Sigma)$$

  - Linked through covariance structure

$$\Sigma = \begin{pmatrix} \sigma_Y^2 & g(\theta) \\ g'(\theta) & \sigma_F^2 \end{pmatrix}$$

- Inverse probability methods - see Scharfstein *et al*

The JoineR package

- Longitudinal data formatting, visualising and simulation
- Joint model class and plotting function
- Simulating data from joint models
- Transformation model and random effects joint model fitting functions

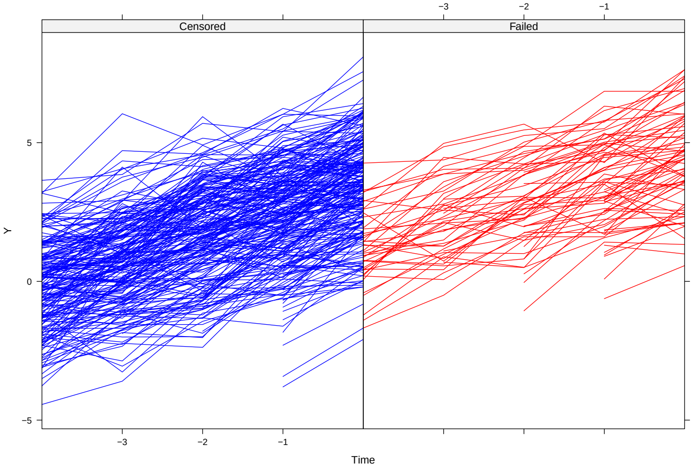
Simulation function

Various simulation studies were carried out to test the software for each possible model. Functions to simulate data are part of the package.

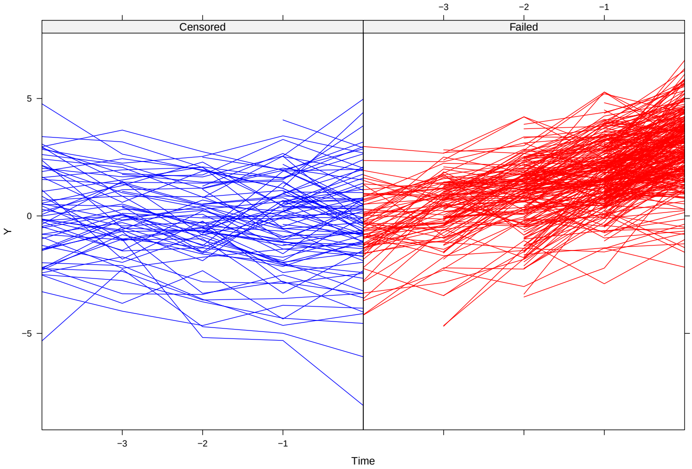
```
sim_intercept <- simjoint(n = 500,model = ``int'',
                          gamma = 3, ntms = 5)
```

- Options for continuous/categorical/factors
- Constant or parametric baseline hazard
- Balanced or unbalanced data
- User can choose level of drop-out/censoring and type of latent association

Plotting simulated data: random intercept model



Plotting simulated data: random intercept and slope model



Simulation study: results for RE model

- Intercept only model: $R_1 = U_0$, $R_2 = \gamma R_1$

n	β_{11}	β_{12}	β_{21}	β_{22}	γ	σ_0^2	σ_ϵ^2
250	1.00	1.00	1.00	1.00	1.01	0.98	0.49
500	1.00	1.00	0.99	0.99	0.98	1.00	0.50
1000	1.00	1.00	1.00	1.00	0.99	1.00	0.50
True	*1*	*1*	*1*	*1*	*1*	*1*	*0.5*

Table: Simulation results from intercept only model

- Intercept and slope models: $R_1 = U_0 + U_1 t$, $R_2 = \gamma R_1$

n	β_{11}	β_{12}	β_{21}	β_{22}	γ	σ_0^2	σ_1^2
250	1.00	0.99	0.99	1.00	0.25	0.99	1.99
500	1.00	0.99	1.01	1.00	0.25	0.99	1.99
1000	1.00	1.00	1.00	1.00	0.25	1.00	2.00
True	*1*	*1*	*1*	*1*	*0.25*	*1*	*2*

Table: Simulation results from intercept and slope model

Application: liver cirrhosis data

Data on almost 500 patients from a randomised clinical trial of prednisone for liver cirrhosis patients. Further details can be found in Andersen *et al.*

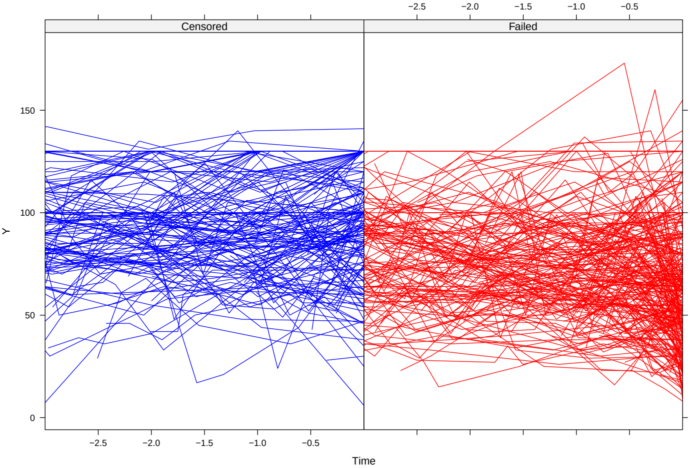
- We can fit a joint model using **JoineR**

```
fit_int_slope <- joint(Y ~ int + P + tt + P_tt + tt0 + P_tt0,
                 ``id'',``tt'', Surv(s,cen)~sP,
                 data = liverJointData,
                 longsep = T, survsep = T)

fit_int_slope <- joint(Y ~ int + P + tt + P_tt + tt0 + P_tt0,
                 ``id'',``tt'', Surv(s,cen)~sP,
                 data = liverJointData,
                 longsep = T, survsep = T, gpt = 15)

fit_quadratic <- joint(Y ~ int + P + tt + P_tt + tt0 + P_tt0,
                 ``id'',``tt'', Surv(s,cen)~sP,
                 data = liverJointData,
                 model = ``quad'' , longsep = T, survsep = T)
```

Liver cirrhosis data



Application: liver cirrhosis data (ctd.)

	Parameter	Estimates	
		Separate analysis	Joint analysis
Longitudinal			
	Intercept	69.99	70.31
	Treatment, P	11.63	11.28
	Time, t	1.33	0.25
	P $\times t$	-1.59	-1.24
	$t = 0$, B	-1.15	-1.48
	$P \times B$	-11.80	-11.45
Survival			
	Treatment	-0.10	-0.08
Association			
	γ	-	-0.04

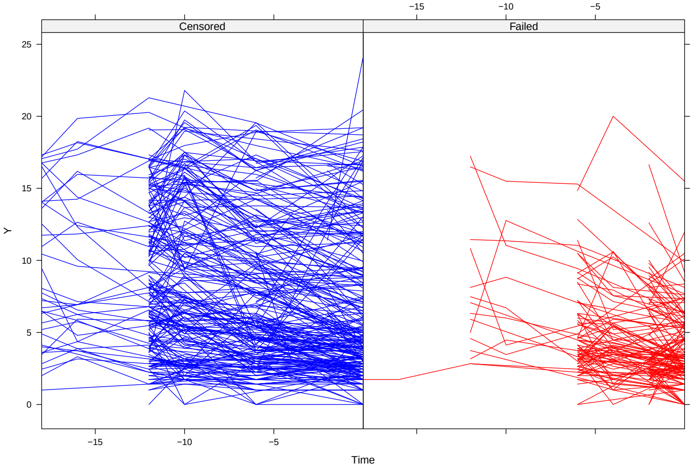
Application II: CD4 cell count data

Data collected on 467 HIV-infected patients to compare efficacy and safety of two antiretroviral drugs. Further details in Guo & Carlin and data available from Brad Carlin's software page.

- We can fit a joint model using **JoineR**

```
fit_int <- joint(Y~ tt + tt_drug + gen + prev + strat,''id'',''tt'',
           Surv(s,cen)~sgrp + sgen + sprev + sstrat, model = ''int'',
           data = CarlinJointData, longsep = T, survsep = T)
```

CD4 cell count data: Guo & Carlin



Application II: CD4 cell count data (ctd.)

	Parameter	Estimates	
		Separate analysis	Joint analysis
Longitudinal			
	Intercept	8.00	7.96
	Time	-0.16	-0.17
	Time $\times$ Drug	0.02	0.02
	Gender	-0.15	-0.12
	Prev OI	-2.31	-2.34
	Stratum	-0.11	-0.14
Survival			
	Drug	0.22	0.30
	Gender	-0.17	-0.17
	Prev OI	0.65	0.65
	Stratum	0.08	0.08
Association			
	γ	-	-0.23

Future work

- Deposit on CRAN
- Added flexibility for latent structure in model fitting - user can choose D_1, D_2
- More flexibility in simulation routines
- See the project website at http://www.liv.ac.uk/joine-r/index.html

Wulfsohn, M. S. & Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53, 330–339.

Henderson, R. , Diggle, P. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1, 465–480.

Diggle, P., Sousa, I. and Chetwynd, A. G. (2007). Joint modelling of repeated measurements and time-to-event outcomes. The fourth Armitage lecture. *Statistics in Medicine*, 27, 2981–2998.

Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1998). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *JASA*, 94, 1096-1146.

Guo, X. & Carlin, B. (2004). Separate and joint modelling of longitudinal and time-to-event data using standard computer packages. *The American Statistician*, 58, 16–24.

Andersen, P. K., Borgan, O, Gill, R. D. & Kieding, N. *Statistical Models based on Counting Processes*. Springer: Berlin, 1997.