Marriage Survival

Gilbert Ritschard

Department of Econometrics and Laboratory of Demography, University of Geneva
http://mephisto.unige.ch

Workshop on Sequence Analysis, Lund, May 8-9, 2008

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES ÉCONOMIQUES ET SOCIALES
Département d'économétrie

Objective of this presentation

- **Colorize your life courses**
- Preliminary results from the analysis of the retrospective Swiss Household Panel (SHP) survey.
- Focus on visualization of life course data.
- Divorce and de-standardization of life Swiss life courses.

Two broad approaches

- Survival analysis (Event history analysis): Focus on one event (divorce)
  - Which factors influence the hazard rate of experimenting the event?
  - What is the importance of these factors?
- Sequence analysis: sequence describing whole life course.
  - Similarity between pairs of state sequences ($\Rightarrow$ cluster analysis).
  - Typical event pattern.
  - Turbulence and other instability measures of a sequence.

Survival Approaches

- Survival or Event history analysis (Blossfeld and Rohwer, 2002)
  - Focuses on one event.
  - Concerned with duration until event occurs or with hazard of experiencing event.
- Survival curves: Distribution of duration until event occurs

$$S(t) = p(T \geq t) \ .$$

- Hazard models: Regression like models for $S(t, \mathbf{x})$ or hazard $h(t) = p(T = t \mid T \geq t)$

$$h(t, \mathbf{x}) = g\Big(t, \beta_0 + \beta_1 x_1 + \beta_2 x_2(t) + \cdots\Big) \ .$$

SHP biographical retrospective survey

http://www.swisspanel.ch

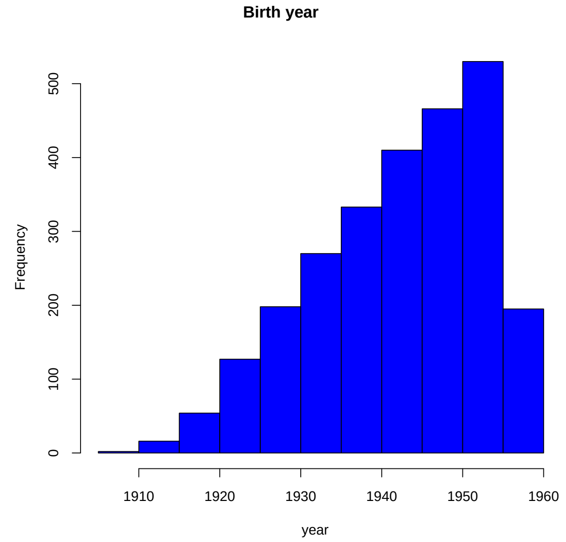
- SHP retrospective survey: 2001 (860) and 2002 (4700 cases).
- We consider only data collected in 2002.
- Data completed with variables from 2002 wave (language).

Characteristics of retained data for divorce

(individuals who get married at least once)

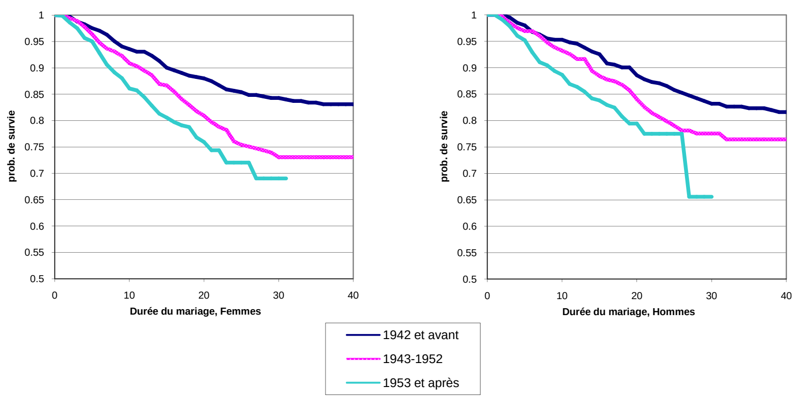
	men	women	Total
Total	1414	1656	3070
1st marriage dissolution	231	308	539
	16.3%	18.6%	17.6%

Distribution by birth cohort



Marriage duration until divorce

Survival curves



Marriage duration until divorce

Hazard model

- Discrete time model (logistic regression on person-year data)
- $\exp(B)$ gives the Odds Ratio, i.e. change in the odd $h/(1-h)$ when covariate increased by 1 unit.

		exp(B)	Sig.
birthyr		1.0088	0.002
university		1.22	0.043
child		0.73	0.000
language	unknwn	1.47	0.000
	French	1.26	0.007
	German	1	ref
	Italian	0.89	0.537
Constant		0.0000000004	0.000

Survival trees: Principle

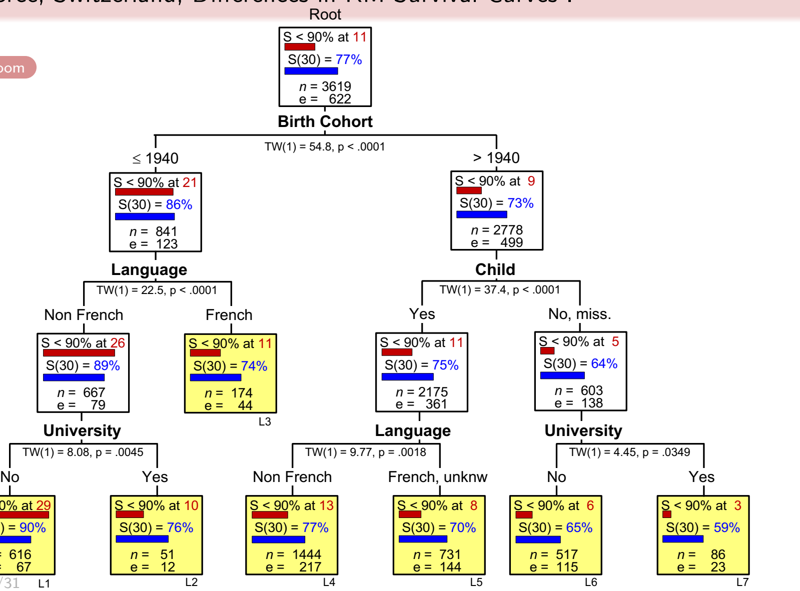
- Target is survival curve or some other survival characteristic.
- Aim: Partition data set into groups that
- differ as much as possible (max between class variability)
  - Example: Segal (1988) maximizes difference in KM survival curves by selecting split with smallest p-value of Tarone-Ware Chi-square statistics

$$TW = \sum_i \frac{w_i\left(d_{i1} - \mathrm{E}(D_i)\right)}{\left(w_i^2 \operatorname{var}(D_i)\right)^{1/2}}$$

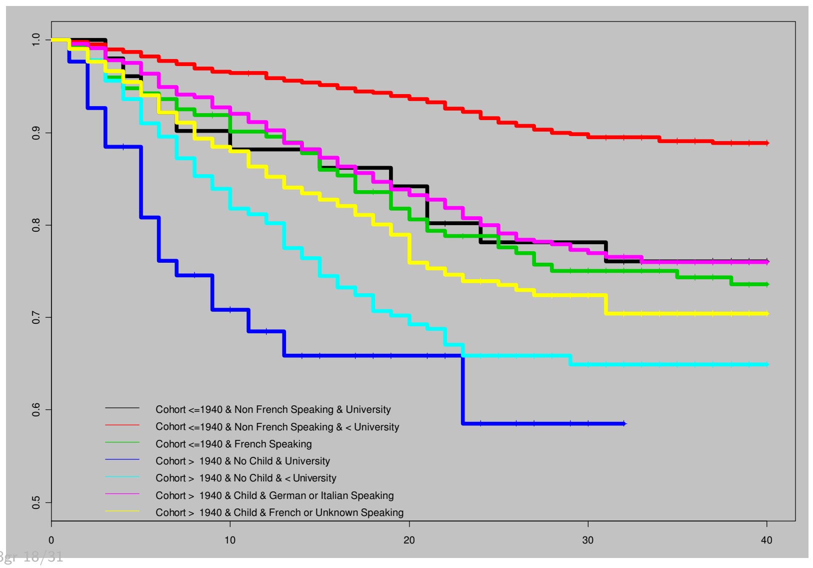
- are as homogeneous as possible (min within class variability)
  - Example: Leblanc and Crowley (1992) maximize gain in deviance (-log-likelihood) of relative risk estimates.

Divorce, Switzerland, Differences in KM Survival Curves I

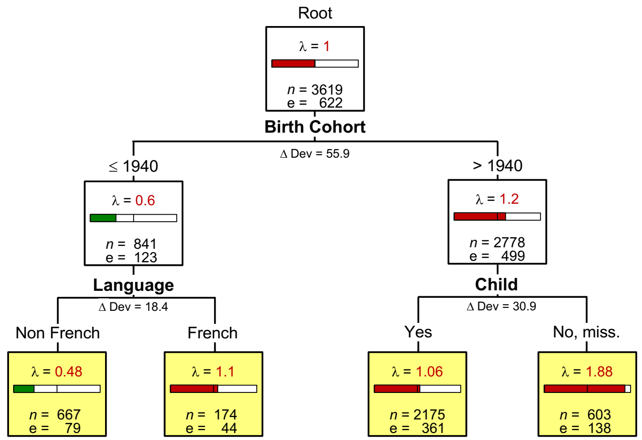
Zoom



Divorce, Switzerland, Differences in KM Survival Curves II



Divorce, Switzerland, Relative risk



Hazard model with interaction

- Adding interaction effects detected with the tree approach
- improves significantly the fit (sig $\Delta\chi^2 = 0.004$)

		exp(B)	Sig.
born after 1940		1.78	0.000
university		1.22	0.049
child		0.94	0.619
language	unknwn	1.50	0.000
	French	1.12	0.282
	German	1	ref
	Italian	0.92	0.677
b_before_40*French		1.46	0.028
b_after_40*child		0.68	0.010
Constant		0.008	0.000

Issues with survival trees in social sciences

1. Dealing with time varying predictors
   - Segal (1992) discusses few possibilities, none being really satisfactory.
   - Huang et al. (1998) propose a piecewise constant approach suitable for discrete variables and limited number of changes.
   - Room for development ...
2. Multi-level analysis
   - How can we account for multi-level effects in survival trees, and more generally in trees?
   - Conjecture: Should be possible to include unobserved shared effect in deviance-based splitting criteria.

Creating duration and censor variables from sequences

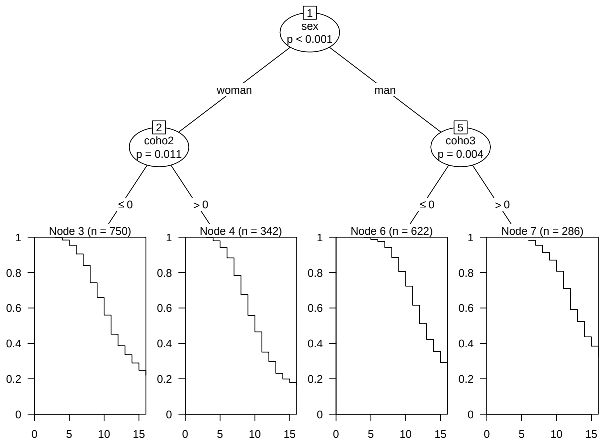
- Duration until marriage.
- The `seqfpos()` function of `TraMineR` returns first occurrence of state.

```
data(biofam)
svar <- 10:25
durmax <- length(svar)

# seqfpos returns the position of first occurrence of the provided state
#   data,s:e  means that we consider the sequence defined in data between columns s and e
#   states considered are
#       2 (married without leaving home)
#       3 (married and leaved home)
#       6 (married with child)
#       7 (divorced)
#   If divorce occurs before any marriage, we assume marriage and divorce the same year
fmar <- data.frame(s2=seqfpos(biofam,svar,2), s3=seqfpos(biofam,svar,3),
    s6=seqfpos(biofam,svar,6), s7=seqfpos(biofam,svar,7))
# creating duration variable as min value of the 4 states
fmar <- data.frame(fmar,fpos=apply(fmar,1,min,na.rm=TRUE))
# create the censor variable mar
fmar <- data.frame(fmar,mar=(fmar$fpos!=Inf))

# Setting duration to sequence length for censored cases.
fmar$fpos[fmar$fpos==Inf] <- durmax
```


Growing a survival tree with `party`

```
## Creting the survival object
library(survival)
surv.fmar <- Surv(fmar$fpos,fmar$mar)

# covariate data frame

coho1 <- (biofam$birthyr < 1940)*"<1940"
coho2 <- (biofam$birthyr >=1940 & biofam$birthyr < 1950)
coho3 <- (biofam$birthyr >=1950)
coho = coho1 + 2*coho2 + 3*coho3
lang <- biofam$plingu02
sex <- biofam$sex

covariates <-data.frame(sex,lang,coho1,coho2,coho3)

library(party)
stree <- ctree(surv.fmar ~ .,data=covariates)
plot(stree,legend)
```

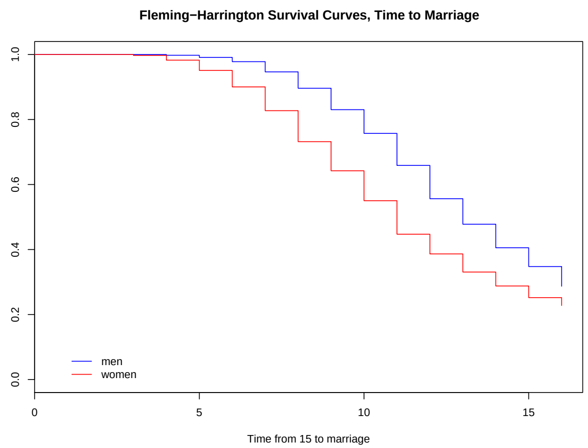
Obtained survival tree for time to marriage



Generating survival curves

```
## Creting the survival object
library(survival)
surv.fmar <- Surv(fmar$fpos,fmar$mar)

surv.fmar <- Surv(fmar$fpos,fmar$mar)

# K-M survival curve

sf.fmar <- survfit(surv.fmar)
summary(sf.fmar)
plot(sf.fmar)

# Fleming-Harrington survival curve

sf.fmar.fh <- survfit(surv.fmar ~ biofam$sex,type="fleming-harrington")
summary(sf.fmar.fh)
plot(sf.fmar.fh,main="Fleming-Harrington Survival Curves, Time to Marriage",
     xlab="Time from 15 to marriage",legend.text=c("men","women"),col=c(4,2))
```

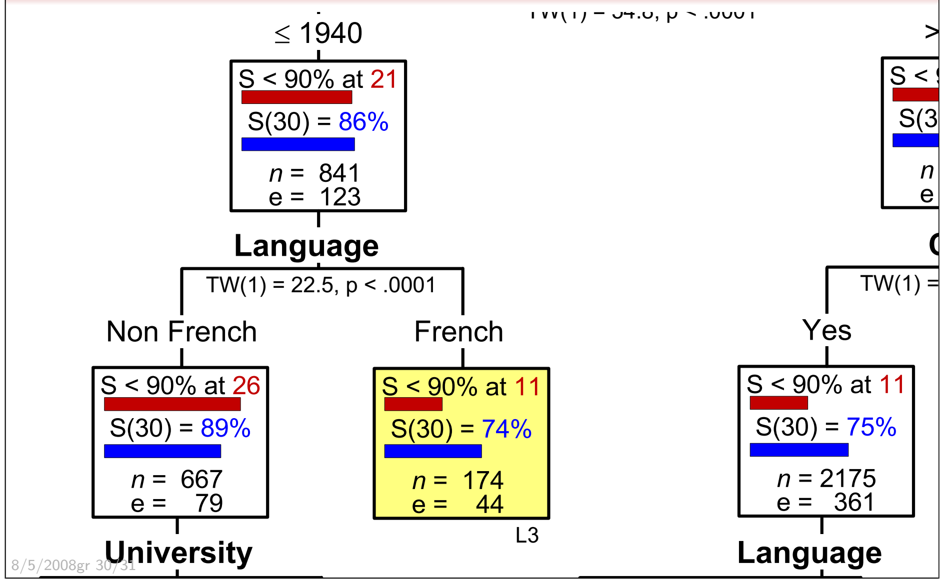

Generated survival curves



THANK YOU!

MERCI !

Divorce, Switzerland, Differences in KM Survival Curves I



References

Blossfeld, H.-P. and G. Rohwer (2002). *Techniques of Event History Modeling, New Approaches to Causal Analysis* (2nd ed.). Mahwah NJ: Lawrence Erlbaum.

Huang, X., S. Chen, and S. Soong (1998). Piecewise exponential survival trees with time-dependent covariates. *Biometrics 54*, 1420–1433.

Leblanc, M. and J. Crowley (1992). Relative risk trees for censored survival data. *Biometrics 48*, 411–425.

Segal, M. R. (1988). Regression trees for censored data. *Biometrics 44*, 35–47.

Segal, M. R. (1992). Tree-structured methods for longitudinal data. *Journal of the American Statistical Association 87*(418), 407–418.