Cohort heterogeneity and competing risks
Statistical Physics Approaches to Systems Biology, Havana, Feb 2019

ACC Coolen
King's College London and Saddle Point Science
Cohort heterogeneity and competing risks
  Regression for time-to-event data
  Consequences and fingerprints
  Informative censoring: the intuition

Bayesian latent class models
  Rationale and definition
  Tests on synthetic data

Applications in cancer research
  Epidemiological cancer data
  Data from failed cancer trials

New directions
  Prospective latent class prediction
Cohort heterogeneity and competing risks
   Regression for time-to-event data
   Consequences and fingerprints
   Informative censoring: the intuition

Bayesian latent class models
   Rationale and definition
   Tests on synthetic data

Applications in cancer research
   Epidemiological cancer data
   Data from failed cancer trials

New directions
   Prospective latent class prediction
Regression for time-to-event data

Data

\[ D = \{(z_1, t_1, r_1), \ldots, (z_N, t_N, r_N)\} \]

\[ z_i = (z_{i1}, \ldots, z_{id}) : \text{d covariates (measured at } t = 0) \]

\[ t_i > 0 : \text{first failure time (death, onset of disease, ...)} \]

\[ r_i \in \{0, 1, \ldots, R\} : \text{failure type (or ‘risk’)} \]

Heterogeneity

visible: variability in the available covariates

latent: variability in host or disease, not visible in the covariates
(individuals with same covariates \( z \) are not clones ...)
Competing risks, identifiability and interpretation

▶ **Competing risks**

Informative censoring, i.e. event times of risks are *statistically dependent*: 
\[ p(t_1, \ldots, t_R | z) \neq \prod_{r=1}^{R} p(t_r | z) \]

reported time: \( t = \min\{t_1, \ldots, t_R\} \)

▶ **Interpretation of crude hazard rates**

Eliminating one risk can change hazard rate of others ...

if hazard rate for risk 1 is low:

(i) event 1 is intrinsically unlikely?
(ii) or it is often preceded by event 2?

to disentangle risks: need \( p(t_1, \ldots, t_R | z) \)

▶ **Tsiatis’ identifiability problem (1975)**

Joint event time distribution \( p(t_1, \ldots, t_R | z) \)
cannot be inferred from survival data alone ...
Conventional methods for analysing time-to-event data

Kaplan-Meier estimators
Cox regression

- not designed to handle disease/host heterogeneity, beyond variability in covariates

- to allow interpretation:
  have to assume different risks are uncorrelated, dangerous when many censoring events ...

random effects models, frailty models, latent class models

- usually constructed for primary risk only, so still cannot handle correlated risks

- do not exploit the link between latent heterogeneity and competing risks ...
Consequences and fingerprints of latent heterogeneity

- **Violation of proportional hazards assumption**

- **Interpretation of time dependencies tricky**

  even if all *individual* hazard rates $h_i$ are time-independent, cohort hazard rate will be time-dependent:

  $$h(t) = \frac{\sum_{i=1}^{n} h_i e^{-h_i t}}{\sum_{i=1}^{n} e^{-h_i t}}$$

- **Interpreting cause-specific survival curves (KM, Cox) no longer possible ...**
If in interpreting our data we assume censoring risks uncorrelated with primary risk
censoring by competing risks can give nonsensical results ...
  – harmful drugs look beneficial
  – beneficial drugs look harmful
  – false protectivity of covariates

would we have spotted this problem if the covariate represented the expression of a specific gene?
Link between cohort heterogeneity and informative censoring

Say 1000 people, two risks, hazard rates $h_A$ and $h_B$

- homogeneous cohort: all individuals have $(h_A, h_B)$

- heterogeneous cohort, but non-informative censoring

\[
\begin{align*}
&\text{homogeneous cohort:} \\
&\text{all individuals have } (h_A, h_B) \\
\end{align*}
\]
Heterogeneity and informative censoring

Say 1000 people, two risks, hazard rates $h_A$ and $h_B$

- homogeneous cohort: all *individuals* have $(h_A, h_B)$

- heterogeneous cohort, *informative cohort filtering*

result: *underestimation of $h_A$*
Heterogeneity and informative censoring

Say 28 people, binary covariate: \( z = 0, 1 \)

associations risk A: \( \beta_A \)

associations risk B: \( \beta_B \)

(B: competing risk, strong)

without risk B:

as many A deaths with \( z = 0 \) as for \( z = 1 \),

overall association \( \beta_A = 0 \)
Heterogeneity and informative censoring

Say 28 people, binary covariate: $z = 0, 1$

associations risk A: $\beta_A$
associations risk B: $\beta_B$
(B: competing risk, strong)

Effect of risk B:

what will we now observe for risk A?
Heterogeneity and informative censoring

what will we now observe for risk A?

A survivors with $z=0$: 6
A survivors with $z=1$: 3
overall association $\beta_A > 0$,
false aetiology
Cohort heterogeneity and competing risks
  Regression for time-to-event data
  Consequences and fingerprints
  Informative censoring: the intuition

Bayesian latent class models
  Rationale and definition
  Tests on synthetic data

Applications in cancer research
  Epidemiological cancer data
  Data from failed cancer trials

New directions
  Prospective latent class prediction
Bayesian latent class methods: rationale and definition

- model all risks simultaneously
- individuals with same covariates can have distinct associations and distinct base hazard rates
- risks are assumed independent only at the level of individuals (this removes Tsiatis’ identifiability problem)
- competing risks, informative censoring: reflect correlated association parameters of different risks

\[
\text{class 1} \\
\text{fraction: } w_1 \\
\text{for all risks } r: \\
h_r^1(t) = \lambda_r^1(t)e^{\beta_r^{11}z_1^1 + \ldots + \beta_r^{1d}z_d^1}
\]

\[
\text{class } L \\
\text{fraction: } w_L \\
\text{for all risks } r: \\
h_r^L(t) = \lambda_r^L(t)e^{\beta_r^{L1}z_1^1 + \ldots + \beta_r^{Ld}z_d^1}
\]

proportional hazards within classes ⇔ proportional hazards at cohort level
independent risks within classes ⇔ independent risks at cohort level
Bayesian analysis and model selection: reliable error bars, and multiple classes only if data demand it

reduces to standard Cox regression if no heterogeneity (Occam’s Razor action of Bayesian model selection)

non-primary events all contribute to latent class inference

fully transparent interpretation, unlike some other competing risk approaches ...

formulae for survival curves decontaminated for informative censoring, and retrospective class allocation of individuals

\[(Rowley \ et \ al, \ SIM, \ 2017)\]
Technicalities ...

- **censoring**
  modelled as ‘risk’ $r = 0$ with no associations

- **data likelihood**

  $$p(t, r|z) = \sum_{\ell=1}^{L} w_{\ell} p(t, r|z, \ell), \quad p(t, r|z, \ell) = \lambda_{\ell}(t)e^{\beta_{\ell} \cdot z - \Lambda_{0}(t) - \sum_{r'=1}^{R} \exp(\beta_{r'} \cdot z) \Lambda_{r'}(t)}$$

- **base rates**
  spline construction for $\{\lambda_{\ell}(t)\}$, with $K$ spline points

- **Bayesian model selection**
  $K$: baserate complexity
  $L$: number of latent classes
  $M$: heterogeneity complexity

- **numerical implementation**
  curvature estimation near parameter boundaries ...
  avoiding local minima in high-dim searches ...
  CPU efficiency ...
Upon determining parameters and hyper-parameters explicit formulae for e.g.

- **covariate-conditioned survival curves and hazard rates:**

  \[
  h_r(t|z) = \frac{\sum_{\ell} w_{\ell} \lambda_{\ell}^r(t) e^{\beta_{r}^\ell \cdot z - \sum_{r'=1}^R \exp(\beta_{r'}^\ell \cdot z) \Lambda_{r'}^\ell(t)}}{\sum_{\ell} w_{\ell} e^{-\sum_{r'=1}^R \exp(\beta_{r'}^\ell \cdot z) \Lambda_{r'}^\ell(t)}} ,
  \]

  **decontaminated:**

  \[
  \tilde{h}_r(t|z) = \frac{\sum_{\ell} w_{\ell} \lambda_{\ell}^r(t) e^{\beta_{r}^\ell \cdot z - \exp(\hat{\beta}_{r}^\ell \cdot z) \Lambda_{r}^\ell(t)}}{\sum_{\ell} w_{\ell} e^{-\exp(\beta_{r}^\ell \cdot z) \Lambda_{r}^\ell(t)}} .
  \]

- **cause-specific cumulative incidence function:**

  \[
  F_r(t|z) = \int_0^t dt' \ e^{-\Lambda_0(t')} \ \sum_{\ell} w_{\ell} \lambda_{\ell}^r(t') e^{\beta_{r}^\ell \cdot z - \sum_{r'=1}^R \exp(\beta_{r'}^\ell \cdot z) \Lambda_{r'}^\ell(t')} .
  \]

- **class membership probabilities:**

  \[
  p(\ell|t, r, z) = \frac{w_{\ell} p(t, r|z, \ell)}{\sum_{\ell'} w_{\ell'} p(t, r|z, \ell')} .
  \]
Tests on synthetic data

inference of classes and parameters

3 classes:
red, blue, green

\[
\beta^\mu \\
\lambda_1(t) e^{\varphi_1^0}
\]

\(N = 200\)
\(N = 2000\)
\(N = 20000\)
Tests on synthetic data
decontaminating survival curves for informative censoring

$S_1^{KM}$: Kaplan-Meier
$S_1$: crude survival curve
$\tilde{S}_1$: decontaminated curves

red dashed: true survival curves
Cohort heterogeneity and competing risks
  Regression for time-to-event data
  Consequences and fingerprints
  Informative censoring: the intuition

Bayesian latent class models
  Rationale and definition
  Tests on synthetic data

Applications in cancer research
  Epidemiological cancer data
  Data from failed cancer trials

New directions
  Prospective latent class prediction
Prostate cancer data
(ULSAM data base, $n = 2047$)

Cox regression:
smoking is protective against PC

negative association with smoking *only* in *extremely frail* subgroup of patients

red class: high overall frailty
green class: low overall frailty

(Rowley et al, SIM, 2017)
Breast cancer data
(AMORIS data base, $N = 1798$)

Cox regression finds no significant associations
(proportional hazards violated)

red class: predominantly younger women
green class: predominantly older women

(Wulaningsih et al, BMC Cancer 2015)
Applications to failed cancer trials

- **failed clinical trials**
  
often some drug benefit, but not enough in view of costs ...
  (in the absence of a biomarker to select patients)

- **two possibilities**
  
  1. there exist measurable differences between individuals that explain response variation, we just don’t know what they are ...
     
     subgroups with distinct quantitative characteristics, cohort is in principle **stratifiable**
  
  2. there are no measurable differences between individuals to explain response variation: cohort **not stratifiable**

- **Bayesian Latent class analysis**
  
  – rational method for determining whether cohort is stratifiable
  
  – retrospective class assignment: tool for identifying latent classes
Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial

Timothy S Maughan, Richard A Adams, Christopher G Smith, Angela M Meade, Matthew T Seymour, Richard H Wilson, Shelley Idziaszczyk, Rebecca Harris, David Fisher, Sarah L Kenny, Edward Kay, Jenna K Mitchell, Ayman Madi, Bharat Jasani, Michelle D James, John Bridgewater, M John Kennedy, Bart Claes, Diether Lambrechts, Richard Kaplan, Jeremy P Cheadle, on behalf of the MRC COIN Trial Investigators

Summary

Background In the Medical Research Council (MRC) COIN trial, the epidermal growth factor receptor (EGFR)-targeted antibody cetuximab was added to standard chemotherapy in first-line treatment of advanced colorectal cancer with the aim of assessing effect on overall survival.

Interpretation This trial has not confirmed a benefit of addition of cetuximab to oxaliplatin-based chemotherapy in first-line treatment of patients with advanced colorectal cancer. Cetuximab increases response rate, with no evidence of benefit in progression-free or overall survival in KRAS wild-type patients or even in patients selected by additional mutational analysis of their tumours. The use of cetuximab in combination with oxaliplatin and capcitabine in first-line chemotherapy in patients with widespread metastases cannot be recommended.
The COIN trial (colorectal cancer)

\( n = 398 \)

- two sub-cohorts, with similar base hazard rates, but distinct overall frailties and associations.
- method provides retrospective class assignment
- new tools to identify \textit{a priori} the responders to Cetuximab?
The TOPICAL trial (lung cancer)

\( n = 580 \)
Cohort heterogeneity and competing risks
  Regression for time-to-event data
  Consequences and fingerprints
  Informative censoring: the intuition

Bayesian latent class models
  Rationale and definition
  Tests on synthetic data

Applications in cancer research
  Epidemiological cancer data
  Data from failed cancer trials

New directions
  Prospective latent class prediction
Prospective latent class prediction

If any of the covariates correlate with retrospective class membership:
(e.g. Amoris)

replace

\[ p(t,r|z) = \sum_{\ell=1}^{L} w_\ell \ p(t, r|z, \ell) \]  \[ \rightarrow \]  \[ p(t,r|z) = \sum_{\ell=1}^{L} w_\ell(z) \ p(t, r|z, \ell) \]

▶ suitable parametrisation \( w_\ell(z) \)

▶ prospective class prediction,
i.e. objective data-driven stratification to rescue failed trials

▶ but increasingly complex models,
many parameters: danger of overfitting ...