Survival analysis with Gaussian processes

Aki Vehtari
Aki.Vehtari@aalto.fi

Aalto University
School of Science

Probabilistic Machine Learning Group
http://research.cs.aalto.fi/pml/

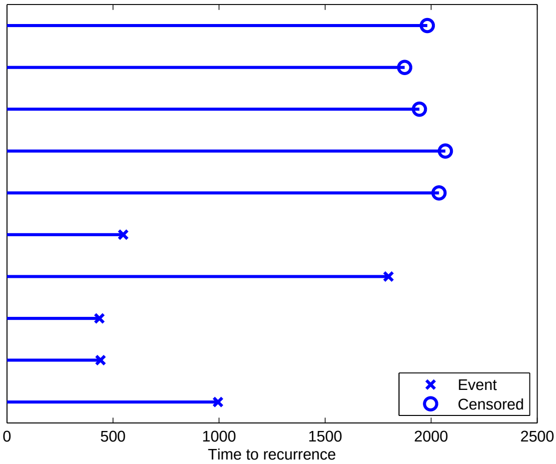
Outline

- Personalised medicine?
- Survival data
- Gaussian processes
- GIST cancer example 1
  - Cox proportional hazards (Cox-PH)
- GIST cancer example 2
  - Time dependent covariates
- Leukemia example
  - Accelerated failure time (AFT) model

Personalised medicine?

- Gastrointestinal stromal tumor (GIST)
  - 400 patients followed after surgery with 1 or 3 year adjuvant imatinib (medicine taken orally) treatment
  - Old question: is 3 year treatment better than 1 year treatment?
  - New question: how long should a specific patient be treated?

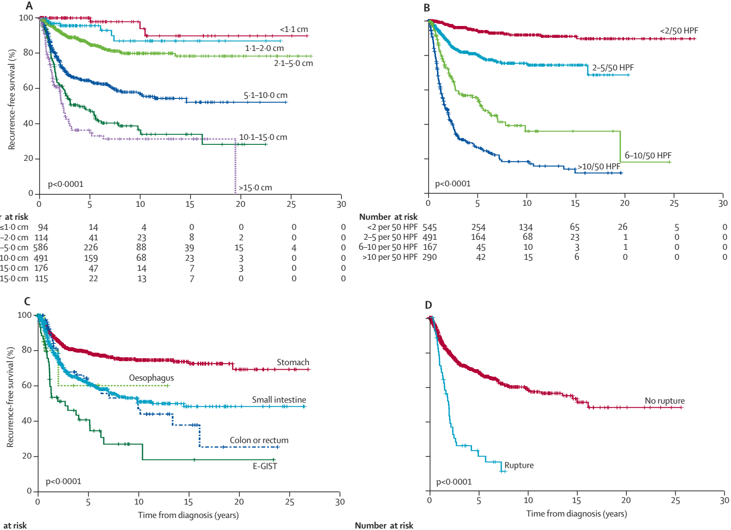
Survival data with time to event



GIST cancer example 1

- Gastrointestinal stromal tumor (GIST)
- 2560 patients followed after surgery (+ 920 validation set)
- Various predictors available

GIST - Kaplan Meier plots



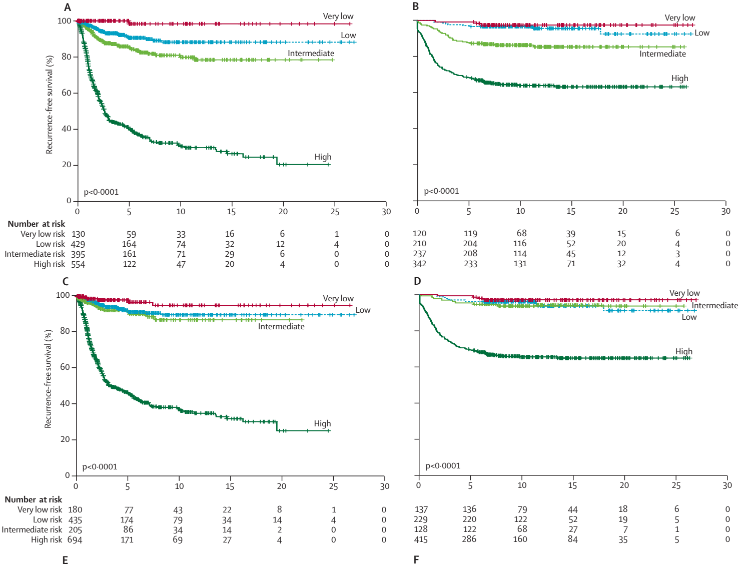
lumber at risk	0	5	10	15	20	25	30
≤1·0 cm	94	14	4	0	0	0	0
1·1–2·0 cm	114	41	23	8	2	0	0
2·1–5·0 cm	586	226	88	39	15	4	0
5·1–10·0 cm	491	159	68	23	3	0	0
10·1–15·0 cm	176	47	14	7	3	0	0
>15·0 cm	115	22	13	7	0	0	0

Number at risk	0	5	10	15	20	25	30
<2 per 50 HPF	545	254	134	65	26	5	0
2–5 per 50 HPF	491	164	68	23	1	0	0
6–10 per 50 HPF	167	45	10	3	1	0	0
>10 per 50 HPF	290	42	15	6	0	0	0

umber at risk	0	5	10	15	20	25	30
Stomach	956	345	146	64	19	3	0
mall intestine	515	162	70	26	8	2	0
Oesophagus	8	2	2	0	0	0	0
olon or rectum	84	32	15	6	1	0	0
E-GIST	61	13	3	2	1	0	0

Number at risk	0	5	10	15	20	25	30
No rupture	1004	270	81	25	6	1	0
Rupture	57	6	0	0	0	0	0

GIST - Kaplan Meier plots



Cox proportional hazards (PH) model

- Hazard (event rate at time t conditional on $T > t$)

$$h(t) = \frac{f(t)}{1 - F(t)}$$

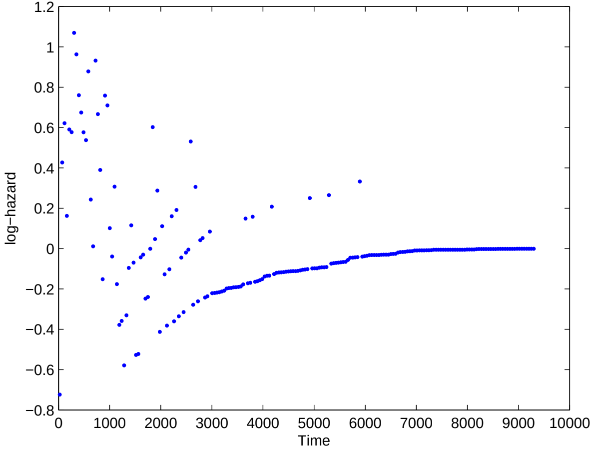
- Proportional hazard model

$$h_i(t) = h_0(t) \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}),$$

  where $h_0(t)$ is baseline hazard in time
  - shape of $h_i(t)$ is determined by $h_0(t)$ and covariates effect only the level of the hazard
- In original Cox PH model, no model for the baseline hazard in time
  - although Cox wrote that such model would improve the hazard estimate

GIST example 1

Baseline log hazard without smoothing

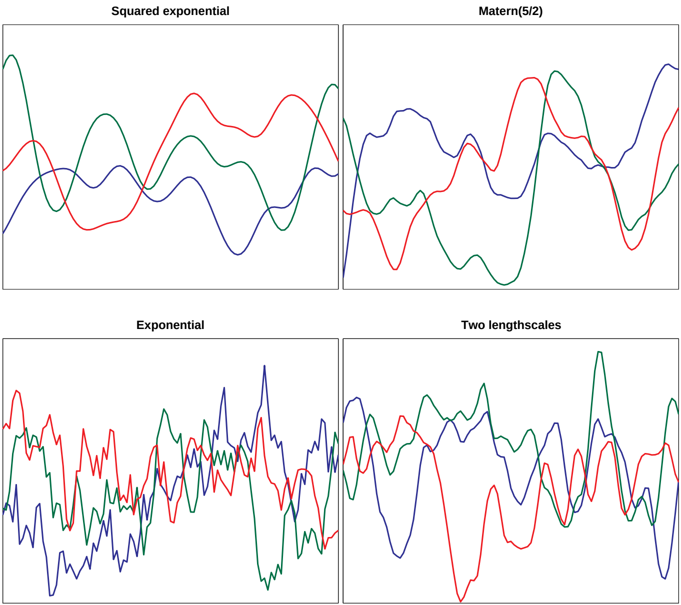


Cox proportional hazards (PH) model

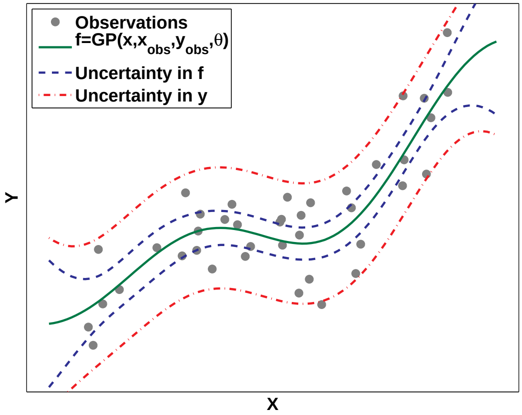
- Alternative to this?

$$h_i(t) = h_0(t) \exp(\boldsymbol{x}_i^T \beta)$$

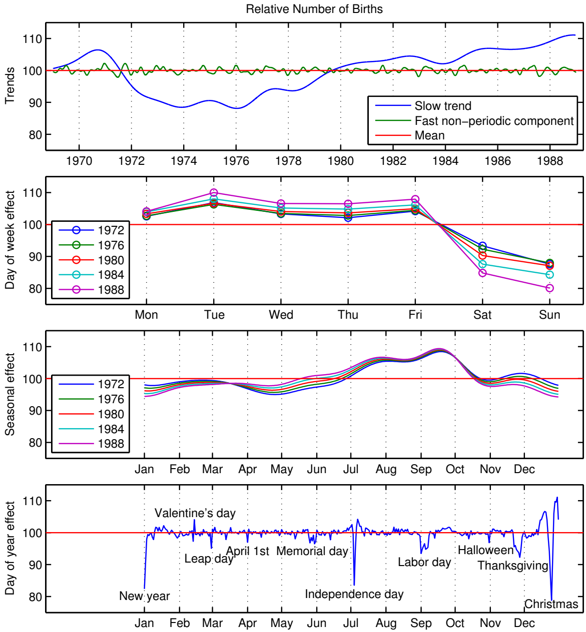
Gaussian process as prior on function space



Posterior Gaussian process given data



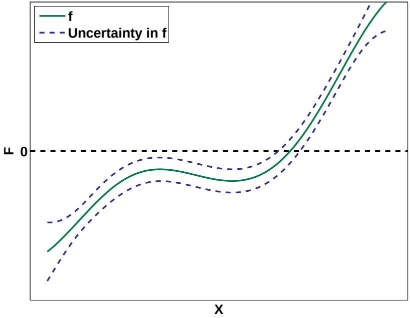
Flexibility of Gaussian process

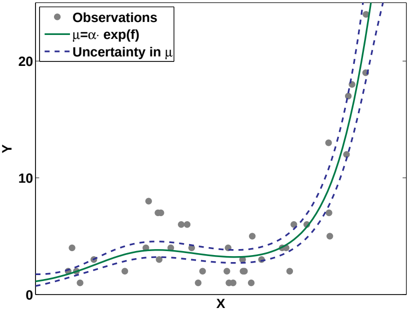


See details in Bayesian Data Analysis, 3rd ed

Non-Gaussian observation models

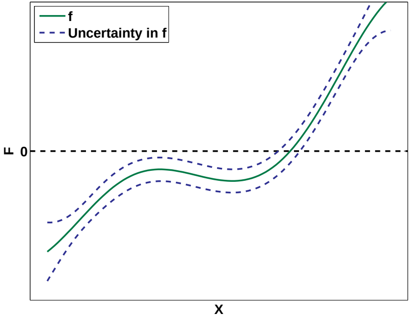
- With a latent variable approach
- E.g. count observations $y \sim \text{Poisson}(\alpha \exp(f))$

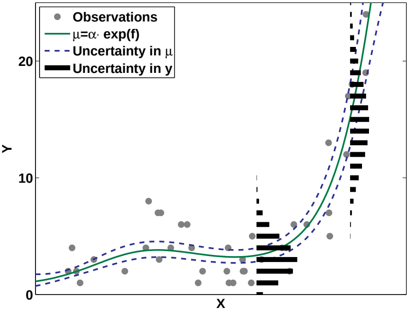




Non-Gaussian observation models

- With a latent variable approach
- E.g. count observations $y \sim \text{Poisson}(\alpha \exp(f))$





GP Cox PH model

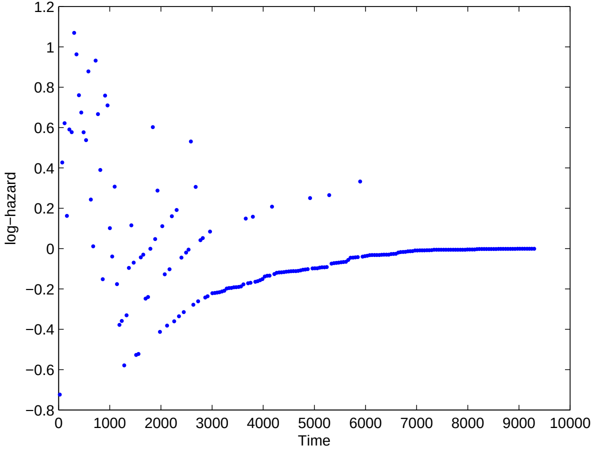
- GP Cox proportional hazard model

$$h_i(t) = \exp(\log(h_0(t)) + \eta_i(\boldsymbol{x}_i)),$$

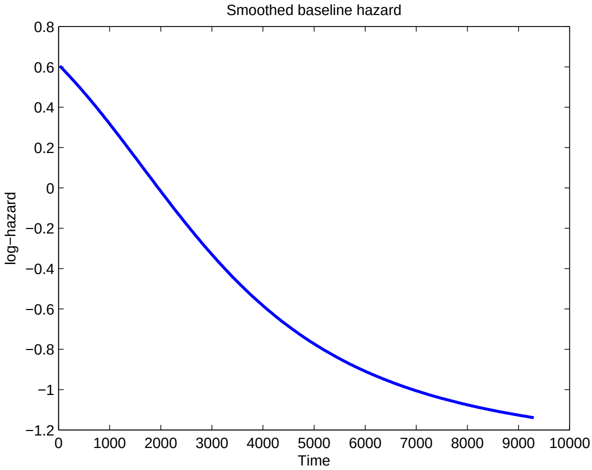
  where $\log(h_0(t))$ and $\eta(\boldsymbol{x}))$ are modeled with GP

- Assuming a piecewise log-constant baseline hazard by partitioning the time axis into K intervals

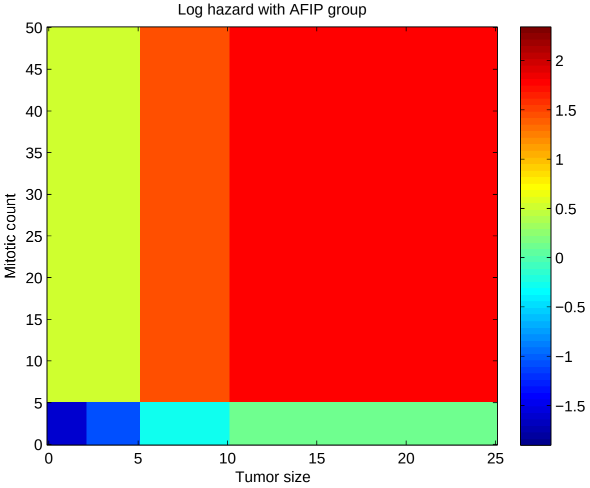
Baseline log hazard without smoothing



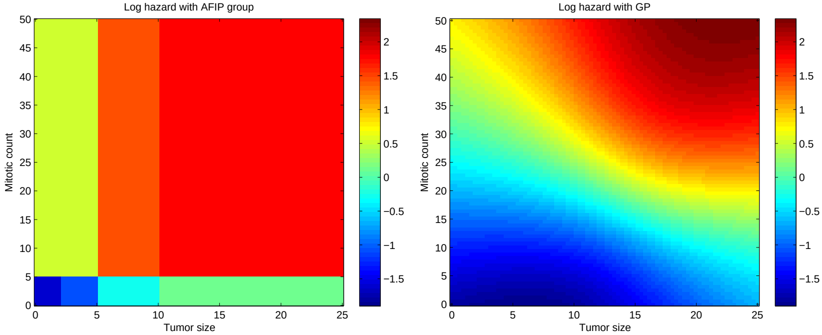
Baseline log hazard with GP



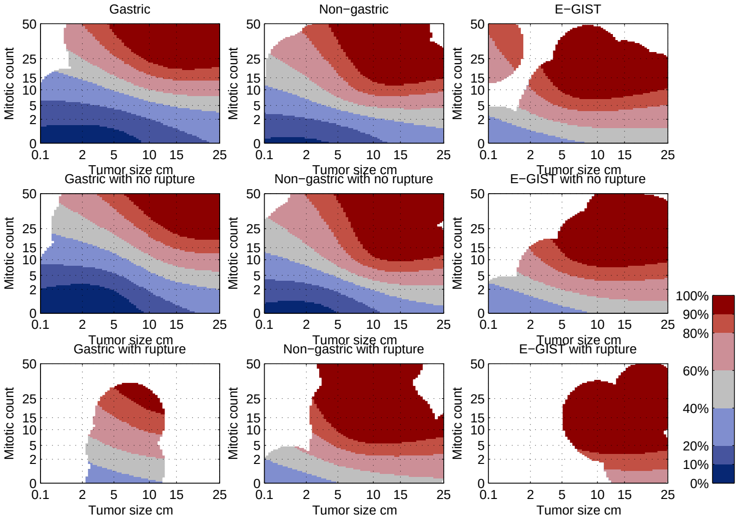
Previously used AFIP risk categorization



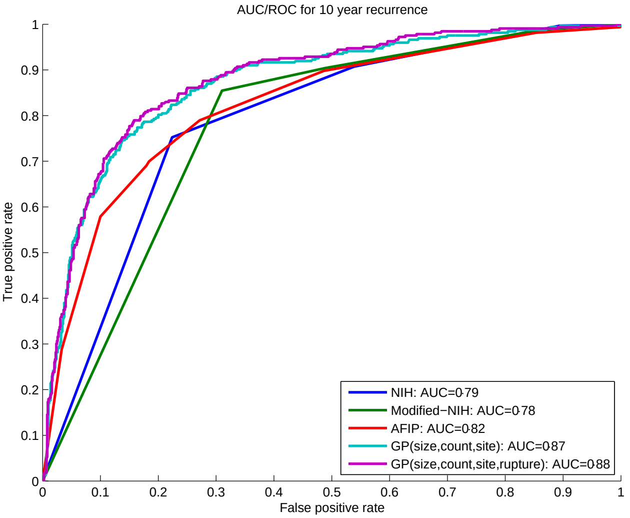
Previously used AFIP risk categorization, and prediction with GP



Probability of recurrence with GP



GIST example 1



GIST example 1

- Joensuu, H., Vehtari, A., Riihimäki, J., Nishida, T., Steigen, S.E., Brabec, P., Plank, L., Nilsson, B., Cirilli, C., Braconi, C., Bordoni, A., Magnusson, M.K., Linke, Z., Sufliarsky, J., Massimo, F., Jonasson, J.G., Dei Tos, A.P. and Rutkowski, P. (2012). Risk of gastrointestinal stromal tumour recurrence after surgery: an analysis of pooled population-based cohorts. The Lancet Oncology, 13(3):265-274.

GIST cancer example 2

- Gastrointestinal stromal tumor (GIST)
- 400 patients followed after surgery with 1 or 3 year adjuvant imatinib (medicine taken orally) treatment
- Covariates
  - the time from the date of randomisation
  - the time from the date of completion of adjuvant imatinib (considered to be 0 before completion of adjuvant therapy)
  - an indicator variable for adjuvant imatinib (considered to be 1 before completion of adjuvant imatinib, and 0 after its completion)
  - tumour mitotic count
  - tumour location (gastric vs. non-gastric)

GIST cancer example 2

- Interval censored data
  - time from the last recurrent free CT scan to next varies
  - CT scanning times are clustered because the scanning intervals are approximately multiplies of 3 months
- Impute unknown recurrence times via sampling

Non-proportional models

- GP non-proportional hazard model

$$h_i(t) = \exp(\eta_i(t, \boldsymbol{x}_i)),$$

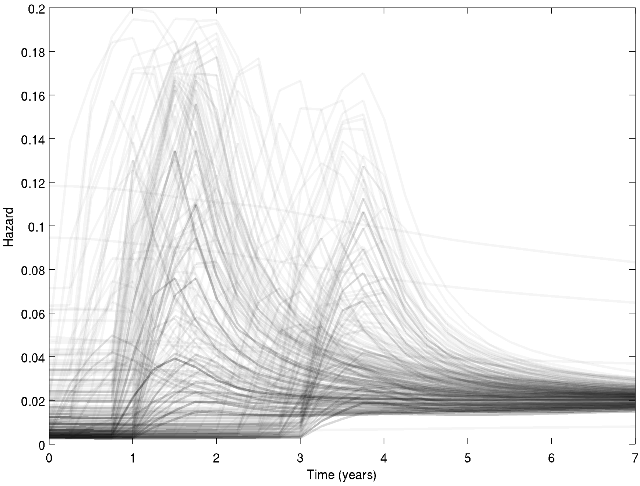
  where $\eta(t, \boldsymbol{x}))$ is modeled with GP

- Assuming a piecewise log-constant baseline hazard
  - we can change to a simpler model...

Non-proportional models – count observation view

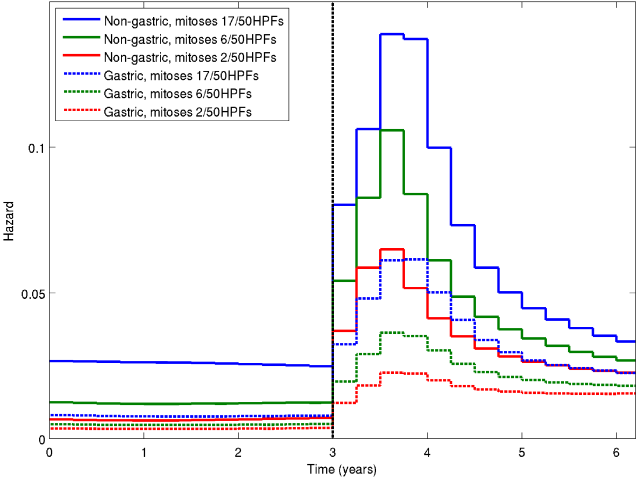
- Assuming discretised time
  - number of events and suspectible people on given time interval with common covariate values
  - observations can be modelled with Poisson/Binomial/Bernoulli model with Gaussian process on location parameter

GIST cancer example 2



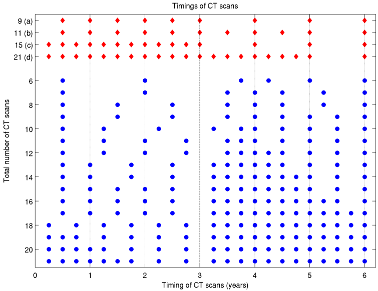
Individual patient hazard functions for GIST recurrence

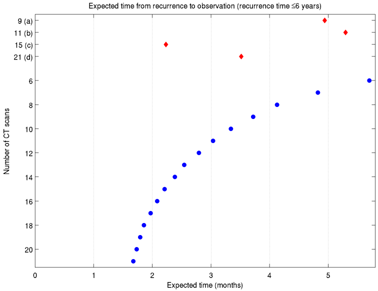
GIST cancer example 2



The hazard of GIST recurrence for six prototype patients

GIST cancer example 2





Left: Timing of CT scans according to a common recommendation and optimised ones.
Right: The expected times to detect the recurrence.

GIST cancer example 2

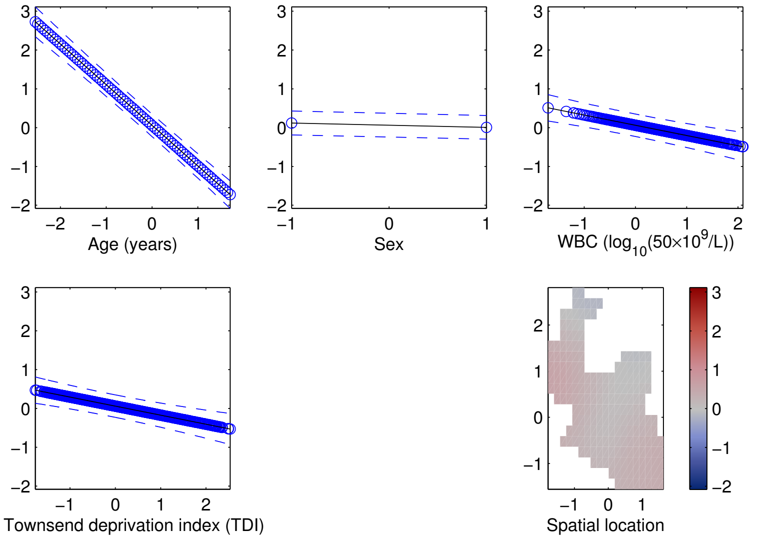
- The numbers of CT scans can be reduced approximately 30% during the first six years of follow-up since initiation of treatment compared with the current follow-up recommendations without jeopardizing early detection of recurrence
- The method may be applicable to the follow-up of other types of human cancer to facilitate early detection of recurrence or to reduce the radiation hazards associated with CT scans
- Joensuu, H., Reichardt, P., Eriksson, M., Hall, K.S., Vehtari, A. (2013). Gastrointestinal Stromal Tumor: A Method for Optimizing the Timing of CT Scans in the Follow-up of Cancer Patients. Radiology, In press.

Accelerated failure time (AFT) models

- Logarithm of the survival time
  - parametric observation model
  - Weibull, log-Gaussian, log-logistic, ...
  - censored observations with cdfs

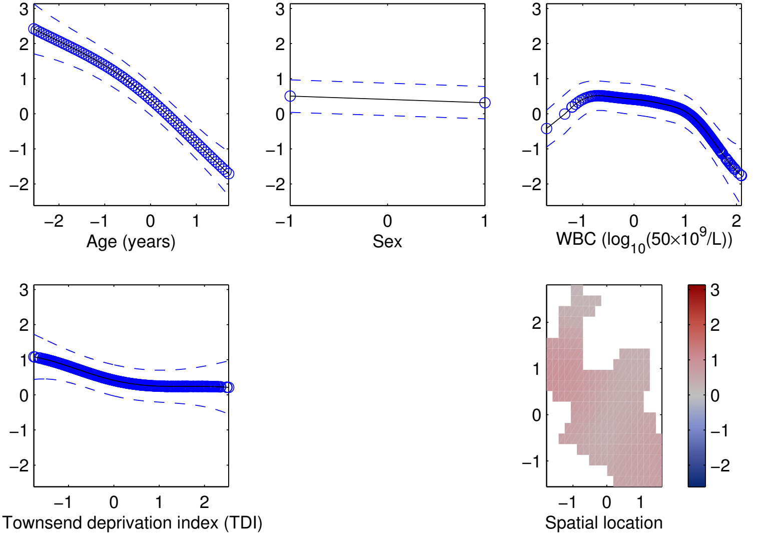
Leukemia survival times

- 1043 cases of acute myeloid leukemia in adults
  - recorded between 1982 and 1998 in the North West Leukemia Register in the United Kingdom

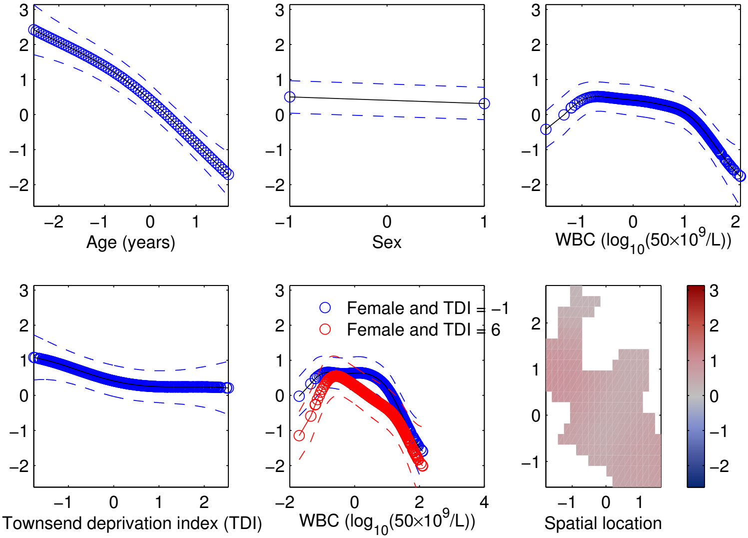


Age (years)
Sex
WBC (log$_{10}$(50×10^9/L))
Townsend deprivation index (TDI)
Spatial location

Leukemia survival times



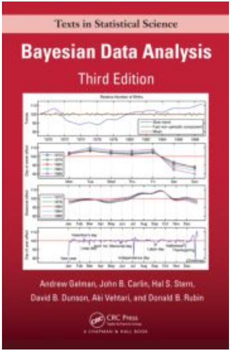
Leukemia survival times



Benefits of GP in AFT models

- Nonlinear effects and implicit interactions (also for more than 2 variables at time)

Leukemia survival times



Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari and Donald B. Rubin (2013). Bayesian Data Analysis, Third Edition. Chapman and Hall/CRC.

- Gaussian process latent variable model (GP-LVM)
  - work in progress
  - joint density model for the data
  - missing data imputation allowing non-Gaussian observation models and non-linear mappings between variables
  - uncertainty in the imputations can be propagated using distributional approximation (instead of sampling)

Computational issues

- n?
  - in paramteric observation models n is the number of patients
  - in non-parameric observation models the effective n is the number of unique patients times the time intervals with suspectible patients
- Computational complexity
  - simple approach $O(n^3)$
  - approximations $O(nm^2)$, where $m \ll n$, may limit the flexibility of the models
  - non-Gaussianity adds extra computation (MCMC, Laplace, EP, VB)

GPstuff

Code for the examples in this presentation available in GPstuff toolbox for Matlab/Octave (with RccpOctave for R)

Jarno Vanhatalo, Jaakko Riihimäki, Jouni Hartikainen, Pasi Jylänki, Ville Tolvanen and Aki Vehtari (2013). GPstuff: Bayesian Modeling with Gaussian Processes. In Journal of Machine Learning Research, 14(Apr):1175-1179.

http://research.cs.aalto.fi/pml/software/gpstuff/

GPy

Some code available soon in GPy toolboxfor Python (by SheffieldML)

http://sheffieldml.github.io/GPy/

Code for Stan in progress (Stan has interfaces in R, Python, Julia, Matlab, Stata)

http://mc-stan.org/

Take home messages

- GPs can be used to improve survival models
  - continuous and discrete valued non-linear functions
  - computation feasible already for many data sets
  - various software tools available
- ... but it's not always easy
  - maybe too slow if using simple generic approaches
  - depends on the software which methods have been implemented
- ... but we are working hard to make this easier!