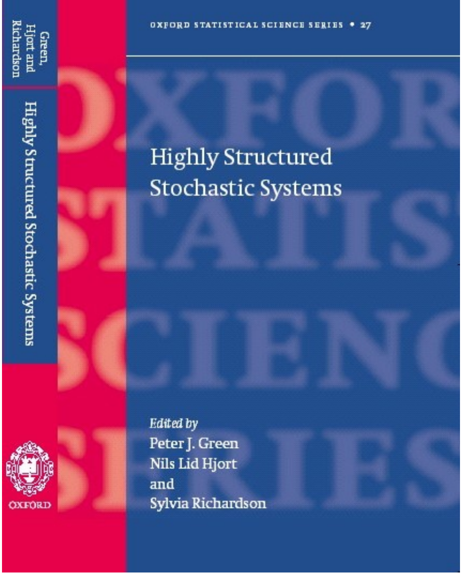
All models are wrong, but some are more plausible than others: cumulative damage processes and time-to-hit models

FOCUS DRIVEN STATISTICAL
INFERENCE WITH COMPLEX DATA

Nils Lid Hjort

Department of Mathematics, University of Oslo

Sylvia's UiO Honorary Doctoral Degree, 31 Aug 2017



HSSS, with European Science Foundation, 1995–2002, Steering Group: Sylvia, E Arjas, A Frigessi, P Green, NLH, S Lauritzen, T O'Hagan, N Wermuth.

General themes

All models are wrong – but some are biologically more plausible than others.

Hope: construction of good models (and then methods) for hazard rates, for survival and event history data, for competing risks, etc.

- Cox model: some non-coherency issues
- Frailty modelling $\Longrightarrow$ classes of hazard rate models
- Bayesian nonparametrics $\Longrightarrow$ classes of hazard rate models
- Cumulative damage process reaches threshold $\Longrightarrow$ models
- Survival as long as all shocks are small $\Longrightarrow$ models
- Parallel damage processes $\Longrightarrow$ competing risks models
- Some damage process never reach threshold $\Longrightarrow$ cure models
- When 'event' is time related $\Longrightarrow$ extended logistic regression

[Footnote: Once upon a time Sylvia and I discussed having a HSSS Research Kitchen on some of these themes. The FocuStat group hosts such a kitchen in November 2017.]

Plan

0 The incoherence of Cox
1 Frailty and cumulative damage processes $\Longrightarrow$ models
2 Time-to-hit models, with gamma processes
3 Applications A & B
4 Beta process jumps $\Longrightarrow$ models
5 Extended logistic regression
6 Competing risks
7 Concluding remarks

0: Issues with Cox type models

Consider survival data with two covariates x_1 and x_2. Cox regression takes hazard to be

$$h(s \mid x_{i,1}, x_{i,2}) = h_0(s) \exp(\beta_1 x_{i,1} + \beta_2 x_{i,2}).$$

There is a model-inconsistency problem here: if we only observe $x_{i,1}$, and calculate the hazard rate $h(s \mid x_{i,1})$, then this will not be of Cox regression form, regardless of distribution of $x_2 \mid x_1$.

Also: if there is perfect Cox structure given x_1 alone, and perfect Cox structure given x_2 alone, one almost never has a Cox model in (x_1, x_2).

Hence: the Cox model suffers from a coherence or plausibility problem. Important: finding good, biologically plausible background explanations that actually imply the Cox structure (or other structures).

1: Frailty processes

There is a broad literature on frailty in survival analysis. These are unobservable or latent explanatory variables accounting for risk-differences between individuals.

In Aalen and NLH (2002):

- some classes of frailty variables, derived via Lévy processes, imply Cox structure;
- some classes of frailty processes also imply Cox structure.

Assume that individual i has covariate x_i and an associated frailty process $Z_i(t)$, growing in time, such that

$$S(t \,|\, x_i, Z_i) = \Pr\{T_i > t \,|\, x_i, Z_i) = \exp\{-Z_i(t)\}.$$

Different models for (the invisible) $Z_i(\cdot)$ lead to different models for

$$S(t \,|\, x_i) = \Pr\{T_i > t \,|\, x_i\} = \mathrm{E}\exp\{-Z_i(t)\}.$$

Cumulative damage processes

Take in particular

$$Z_i(t) = \sum_{j \le M_i(t)} \theta_i G_{i,j},$$

where $M_i(\cdot)$ is a Poisson process with rate $\lambda_i(\cdot)$ and the $G_{i,j}$ are i.i.d., as in cumulative shock model. Then

$$S(t \mid x_i, Z_i) = \prod_{j \le M_i(t)} \exp(-\theta_i G_{i,j}),$$

leading to

$$S(t \mid x_i) = \mathrm{E}\, L_0(\theta_i)^{M_i(t)} = \exp[-\Lambda_i(t)\{1 - L_0(\theta_i)\}].$$

Here $L_0(\theta_i) = \mathrm{E} \exp(-\theta_i G_{i,j})$ is the Laplace transform of the $G_{i,j}$, and $\Lambda_i(t) = \int_0^t \lambda_i(s)\,\mathrm{d}s$.

Different models for $\lambda_i(s)$, for θ_i and $G_{i,j}$, in terms of the covariate x_i, yield hazard rate regression models, via

$$h(s \mid x_i) = \lambda_i(s)\{1 - \mathrm{E} \exp(-\theta_i G_i)\}.$$

Among many possibilities: θ_i constant over individuals; G_i same distribution across individuals; and $\lambda_i(s)$ as in multiplicative Poisson regression, with $\lambda_0(s)\exp(x_i^{\rm t}\beta)$. This frailty process construction then implies the Cox structure:

$$h(s\,|\,x_i) = \lambda_0(s)\exp(x_i^{\rm t}\beta)\{1 - {\rm E}\exp(-\theta G)\}.$$

Competing models also emerge naturally. Among them:

$$h(s\,|\,x_i) = \lambda_0(s)\exp(x_i^{\rm t}\beta)\frac{\exp(x_i^{\rm t}\gamma)}{1+\exp(x_i^{\rm t}\gamma)},$$

e.g. De Blasi and Hjort (2007). Also: additive regression models, via additive model for Poisson rate.

'Twin times' models, via frailty processes

$$Z_0(t) + Z_1(t) \quad \text{and} \quad Z_0(t) + Z_2(t).$$

These have convenient joint Laplace transforms.

2: Time-to-hit models

Time-to-hit-threshold models: Frailty process considerations also inspire non-Cox regression models. Let

$$T_i = \min\{t \geq 0\colon Z_i(t) \geq c_i\}$$

where

$$Z_i(t) \sim \mathrm{Gamma}(aM_i(t), 1) \text{ a Gamma process}.$$

Then

$$\begin{aligned} S_i(t) &= \Pr\{T_i \geq t\} = \Pr\{Z_i(t) < c_i\} \\ &= G(c_i, aM_i(t), 1) = \int_0^{c_i} g(x, aM_i(t), 1)\,\mathrm{d}x. \end{aligned}$$

This is a large class of models, with many shapes for hazards $h_i(t)$, depending on M_i and size of c_i. With acceleration factors $G(c, aM(\kappa_i t))$ we may have crossing hazards.

Thus the nonparametric process view generates fresh regression models (parametric, semiparametric or nonparametric).

One version: as above, with regression on both threshold and acceleration:

$$Z_i(t) \sim \mathrm{Gamma}(M_0(\exp(x_i^{\mathrm{t}}\gamma)t), 1) \quad \text{hits} \quad c_i = \exp(x_i^{\mathrm{t}}\beta).$$

This is parametric if M_0 fixed; may also employ a prior for M_0.

3a: Application A: time to 2nd child after stillbirth

From the Norwegian Birth Registry: 451 married women whose first child died at birth (stillbirth). We read off T, the number of months till the birth of the 2nd child. Model: 2nd child is born when

$$Z(t) \sim \mathrm{Gamma}(aM(t), 1) \geq c,$$

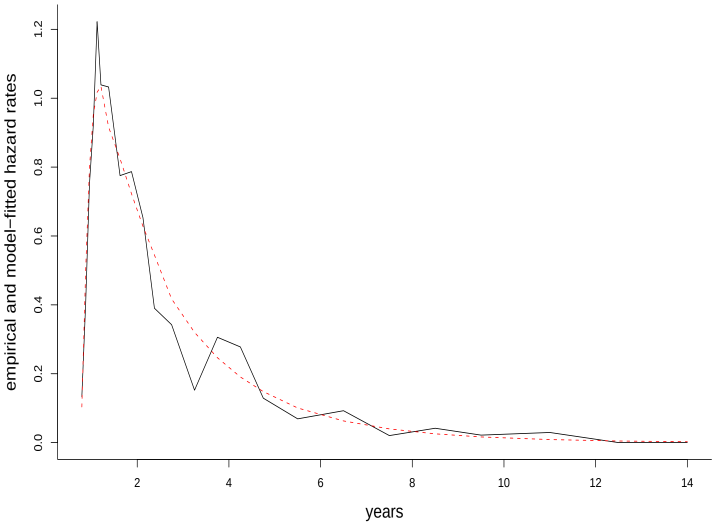
with $M(t) = G(\theta(t - t_0), d, 1)$, and $t_0 = 9/12$ (time in years).

I find ML estimates $(\widehat{a}, \widehat{c}, \widehat{\theta}, \widehat{d})$ from the 451 observations – with a bit of trouble and care, since observations are on interval form, with $\Delta N_j \sim \mathrm{Bin}(Y_j, h_j)$, data for interval $[\ell_j, r_j]$:

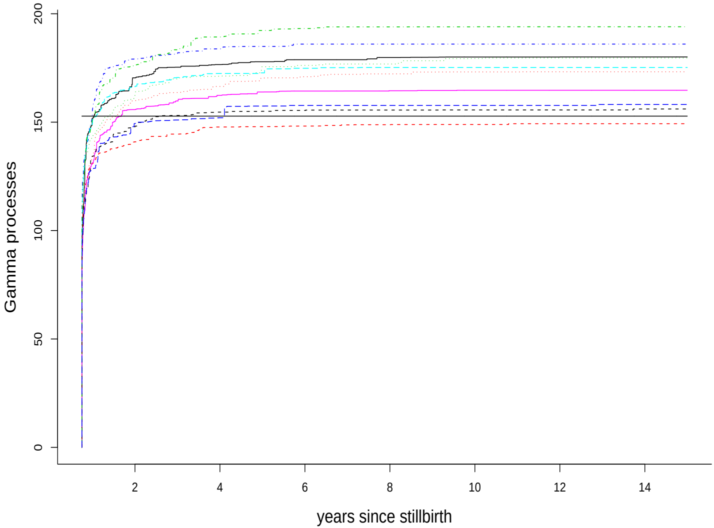
$$h_j = \Pr\{T \in [\ell_j, r_j] \mid T \geq \ell_j\} = 1 - \frac{G(c, aM(r_j), 1)}{G(c, aM(\ell_j), 1)}$$

for the different time intervals.

Model fits very well (better than alternatives), also for the $T = \infty$ individuals; cf. cure models and Stoltenberg's talk.



Empirical and model-fitted hazard rates for the event of a second childbirth, after experiencing a first-born stillbirth, for a population of 451 married Norwegian women.



Simulated Gamma processes for ten women. The process need to cross the level $\widehat{c} = 152.78$ (also plotted in the diagram), in order for a woman to have a 2nd child. With probability $p = G(c, a, 1) \doteq 0.097$, there will never be a 2nd child.

3b: Application B: regression for oropharynx survival data

Survival data $(t_i, \delta_i, x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4})$ for $n = 193$ individuals, with

- x_1: sex (1 male, 2 female);
- x_2: condition (1-2-3-4, index of disability);
- x_3: T-stage (1-2-3-4, size and infiltration of tumour);
- x_4: N-stage (0-1-2-3, index of lymph node metastatis).

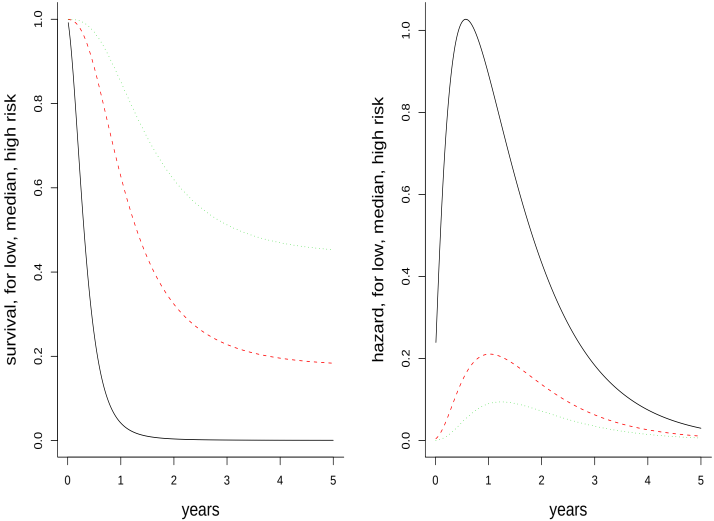
I take the gamma process time-to-hit model

$$t_i = \min\{t \geq 0\colon Z_i(t) \geq c_i\},$$

with $Z_i(t) \sim \mathrm{Gamma}(aM_i(t), 1)$, $M_i(t) = 1 - \exp(-\kappa_i t)$,

$$\begin{aligned} c_i &= \exp(\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_4 x_{i,4}), \\ \kappa_i &= \kappa_0 \exp(\gamma(x_{i,2} + x_{i,3} + x_{i,4})), \end{aligned}$$

with at most $1 + 5 + 2 = 8$ parameters. It does better than Aalen–Gjessing (2001) and other models (in terms of AIC and FIC scores).



Estimated survival curves $S(t)$ and hazard rate functions $h(t)$ are plotted for three individuals, corresponding to high risk ($c = 0.20$), medium risk ($c = 0.65$) and low risk ($c = 0.90$). Hazards are not proportional (so Cox regression does worse).

4: Survival models via Beta process jumps

Consider a Beta process, $A \sim \mathrm{Beta}(c, A_0)$. For a time window $[0, \tau]$ there is a representation

$$A(t) = \sum_{j=1}^{\infty} J_j I\{\xi_j \leq t\} \quad \text{for } t \in [0, \tau]$$

which makes it possible to set up new survival models based on properties of jumps and their locations.

Survival as long as jumps are small: Suppose survival means all jumps smaller than some threshold $x \in (0, 1)$:

$$\Pr\{T > t\} = \Pr\{J_j \leq x \text{ for all jumps on } [0, t]\} = \exp\{-A_0(t) L(x)\},$$

with $L(x)$ from the Lévy representation. This leads to proportional hazards and to Cox regression.

A class of survival distributions emerges by placing distributions on the threshold x.

A special case gives a generalised Cox model:

$$h_i(s) = a_0(s) G(\exp\{-\exp(z_i^{\mathrm{t}}\gamma)\} \,|\, c),$$

where

$$G(x \,|\, c) = \int_x^1 s^{-1} c(1-s)^{c-1}\,\mathrm{d}s.$$

For $c = 1$ we have $-\log x$ and the Cox model.

Survival up to j-th big shock: Further analysis of the Beta process jumps leads to

$$\begin{aligned}\Pr\{T \geq t\} &= \Pr\{\text{the } j \text{ biggest jumps } \leq x \text{ over } [0,t]\} \\ &= \exp\{-A_0(t)L(x)\}\Big[1 + A_0(t)L(x) + \cdots + \frac{\{A_0(t)L(x)\}^j}{j!}\Big].\end{aligned}$$

There are various other cases worth exploring, some of which contain the Cox model as an inner point.

5: Extended logistic regression

Standard logistic regression:

$$p_i = \Pr(Y_i = 1 \,|\, x_i) = \frac{\exp(x_i^{\rm t}\beta)}{1 + \exp(x_i^{\rm t}\beta)}$$
$$= G\big(\log\{1 + \exp(x_i^{\rm t}\beta)\}, 1, 1\big),$$

with $G(\cdot, a, 1)$ the c.d.f. of $\mathrm{Gamma}(a, 1)$.

Extension:

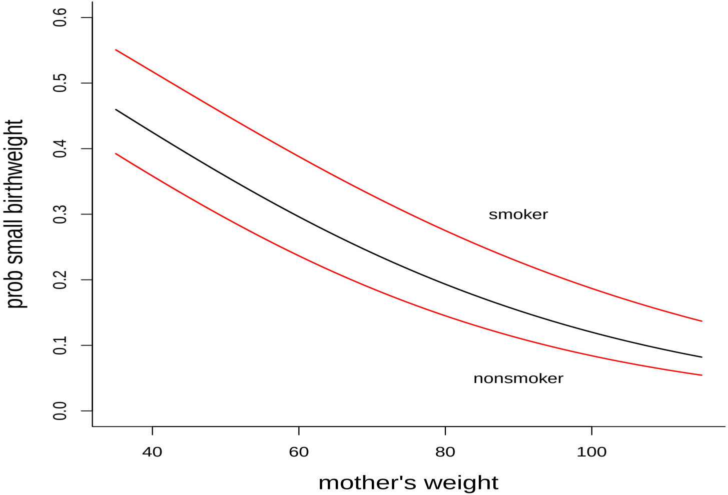
$$p_i = \Pr(Y_i = 1 \,|\, x_i, z_i) = G\big(\log\{1 + \exp(x_i^{\rm t}\beta)\}, a_i, 1\big),$$

where $a_i = a(z_i)$. Could have $a_i = \exp(z_i^{\rm t}\gamma)$, and with some covariates for the x_i part and others for the z_i part.

These models, where 'event' is seen as a gamma process reaching a threshold, are often better than plain logistic regressions, in terms of AIC and FIC scores.

Illustration: probability of child having birthweight ≤ 2.50 kg.
With $x_{i,1}, x_{i,2}$ weight and age of mother,

$$p_i = \begin{cases} G(\log\{1+\exp(\beta_0+\beta_1 x_{i,1}+\beta_2 x_{i,2})\}, 1+\delta, 1) & \text{if smoker} \\ G(\log\{1+\exp(\beta_0+\beta_1 x_{i,1}+\beta_2 x_{i,2})\}, 1-\delta, 1) & \text{if nonsmoker.} \end{cases}$$



6: Competing risks

Suppose each individual has two cumulative risk processes $R_1(t)$ and $R_2(t)$ in his or her rucksack. There is event (e.g. death) when either of these hit threshold c – of cause 1, if R_1 is first; of cause 2, if R_2 is first.

First, new survival models emerge by working with new settings, with $T = \min(T_1, T_2)$, etc. An easy instance is

$$S(t) = \Pr\{T \geq t\} = \{1 - G(c, a_1 M_1(t), 1)\}\{1 - G(c, a_2 M_2(t), 1)\}.$$

Second, can set up models and methods for competing risks. Simple setup:

$$R_j(t) \sim \mathrm{Gamma}(a_j M_j(t), 1) \quad \text{for } j = 1, 2,$$

with independence. Can then estimate all parameters from this type of survival data,

$$(t_i, x_i, \delta_i), \quad \delta_i \in \{0, 1, 2\}.$$

Can also carry out the necessary characterisations and formalisation of likelihood components etc. for the case of

$$R_1(t) = Z_0(t) + Z_1(t), \quad R_2(t) = Z_0(t) + Z_2(t),$$

with independent gamma processes Z_0, Z_1, Z_2 (so full ML analysis is amenable). This opens up for dependent risk processes.

This machinery also leads to formulae for relevant statistical parameters and functions, like

$$q_j(t) = \mathrm{Pr}\{\text{death of cause } j, \text{at } t \,|\, \text{death at time } t\}$$

for $j = 1, 2$. Theory for ML works well enough to supply also confidence bands etc.

See Cunen's talk on WoR and GoT:

$$\mathrm{Pr}\{\text{death by violence at } t \,|\, \text{death at age } t\}.$$

7: Concluding remarks

1. Too often statisticians employ off-the-shelf models and methods.

2. My themes evolve around plausible processes $\Longrightarrow$ good models (and then good methods). Of course there is a literature on such themes (Aalen, Borgan, Gjessing, others), but there is scope for more groundwork.

3. Many of the models coming out of plausible processes are amenable to ML and Bayes analyses etc.; some are semiparametric or nonparametric, with more work to be carried out.

4. Starting with classes of plausible processes one quickly has a plethora of candidate models – so scope for more work, sorting the Very Good Models from the not-as-successful models, e.g. using model selection and model screeing methods (AIC, BIC, FIC).

5. Dynamics can be put into many of the models (covariates changing over time; regime shifts).

6. Models can be individualised, with applications for personalised medicine etc.

(Some) references

Aalen, O.O. and Gjessing, H. (2001). Understanding the shape of the hazard rate: A process point of view. *Statistical Science*.

Aalen, O.O., Borgan, Ø. and Gjessing, H. (2008). *Survival and Event History Analysis: A Process Point of View.* Springer.

Aalen, O.O. and Hjort, N.L. (2002). Frailty models that yield proportional hazards. *Statistics and Probability Letters.*

Claeskens, G. and Hjort, N.L. (2008). *Model Selection and Model Averaging.* Cambridge University Press.

De Blasi, P. and Hjort, N.L. (2007). Bayesian survival analysis in proportional hazard models with logistic relative risk. *Scandinavian Journal of Statistics.*

Gjessing, H., Aalen, O.O. and Hjort, N.L. (2003). Frailty models based on Lévy processes. *Advances in Applied Probability.*

Green, P.J., Hjort, N.L., and Richardson, S. (2003). *Highly Structured Stochastic Systems.* Oxford University Press.

Hjort, N.L., Holmes, C., Müller, P. and Walker, S.G. (2010). *Bayesian Nonparametrics.* Cambridge University Press.

Schweder, T. and Hjort, N.L. (2016). *Confidence, Likelihood, Probability.* Cambridge University Press.