Survival and Event History Models via Gamma Processes

Nils Lid Hjort (with Céline Cunen)

Department of Mathematics, University of Oslo

FocuStat VVV Conference, May 2018

Process to Model:

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

Plan

0 The incoherence of Cox; frailty and cumulative damage processes $\Longrightarrow$ models

1 Gamma process time-to-hit $\Longrightarrow$ models

2 Applications A, B, C

3 Gamma process jumps $\Longrightarrow$ models

4 Extended logistic regression (with brief application)

5 Competing risks (with brief application)

6 Frailtifying the threshold model

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).

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

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

leading to

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

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

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

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

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

Competing models also emerge naturally. Among them:

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

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.

1: 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

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

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:

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

This is parametric if M_0 fixed; may also employ a semi- or nonparametric M_0, or a prior for this function.

$\exists$ links and connections to other work, time-to-hit, threshold regressions, etc., by Aalen, Borgan, Gjessing, Lee, Whitmore, yet others.

(Cf. talks by Emil Stoltenberg and Alex Whitmore.)

2a: Application A: lifelengths in Roman Era Egypt

82 men and 49 women from Egypt, 1st century b.C.; range 0.5 to 96 years.
Gamma process threshold model: men and women die when

$$\begin{aligned} Z_m(t) &\sim \mathrm{Gam}(aM(t), 1) && \geq c, \\ Z_w(t) &\sim \mathrm{Gam}(aM(t) + d\,\mathrm{extra}(t), 1) && \geq c, \end{aligned}$$

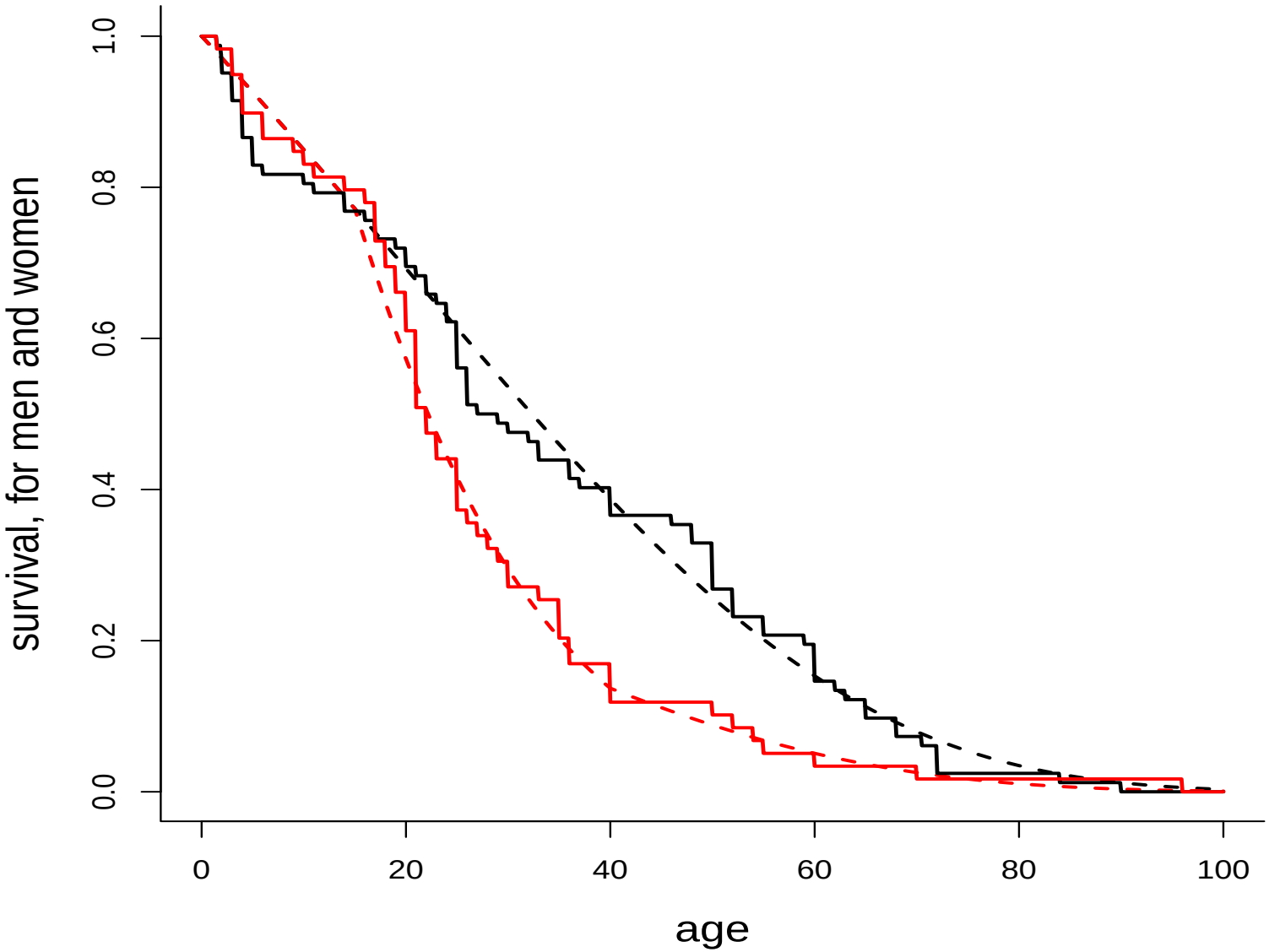
where $M(t) = \exp(\kappa t) - 1$, with the same speed a and same level threshold c for both men and women, and $\mathrm{extra}(t)$ the additional base risk function for being a woman through age window $[15, 40]$.

Can programme and maximise

$$\ell_m(a, \kappa, c) + \ell_w(a, \kappa, c, d).$$

Very good fit to data, better AIC scores than for various other models.

Kaplan–Meier curves along with gamma process based estimated survival curves:



2b: Application B: 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

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

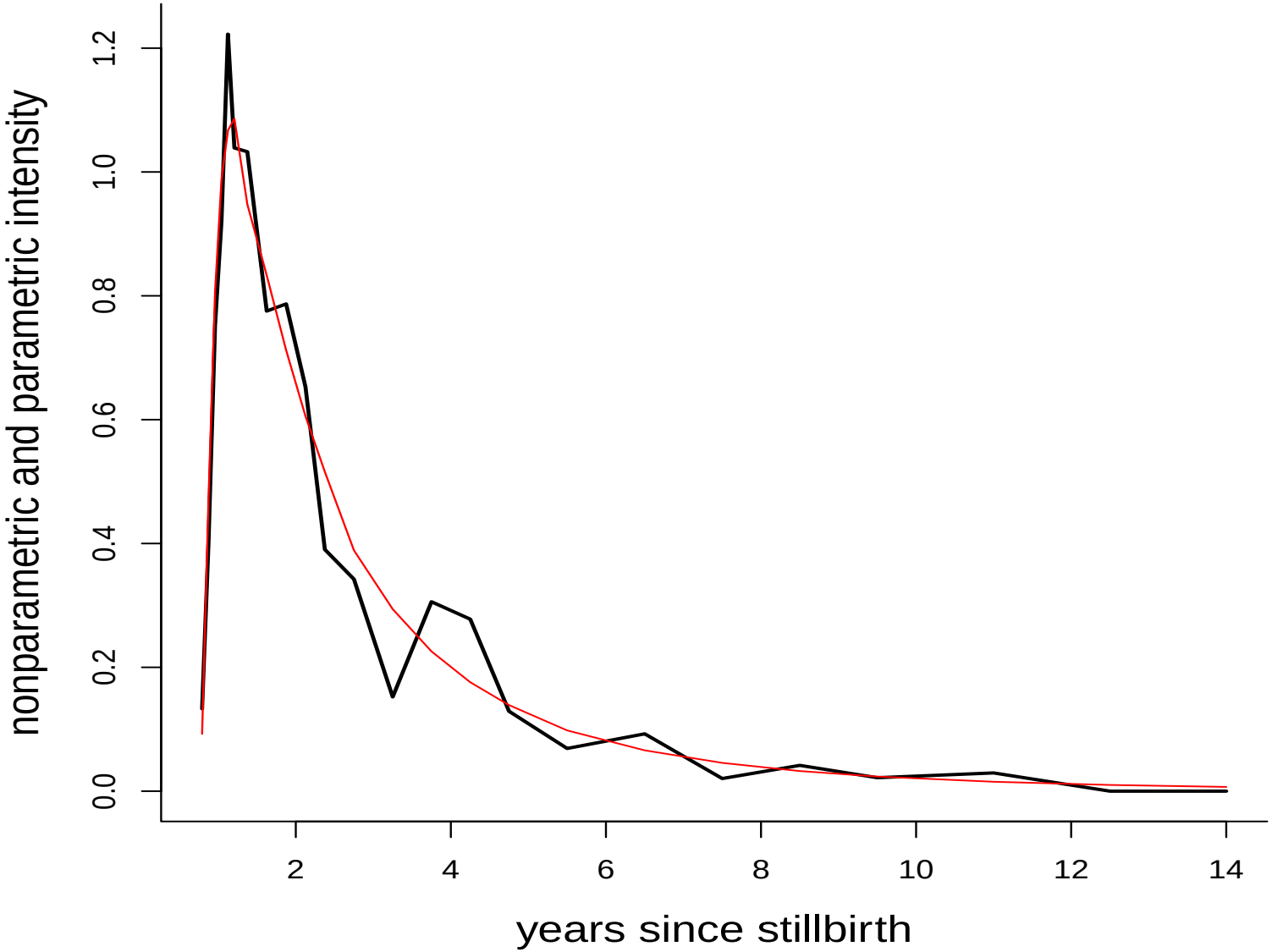
with $M(t) = 1 - \exp[-\{(t - t_0)/\theta\}^d]$, and $t_0 = 9/12$ (time in years).

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 $[\ell_j, r_j]$:

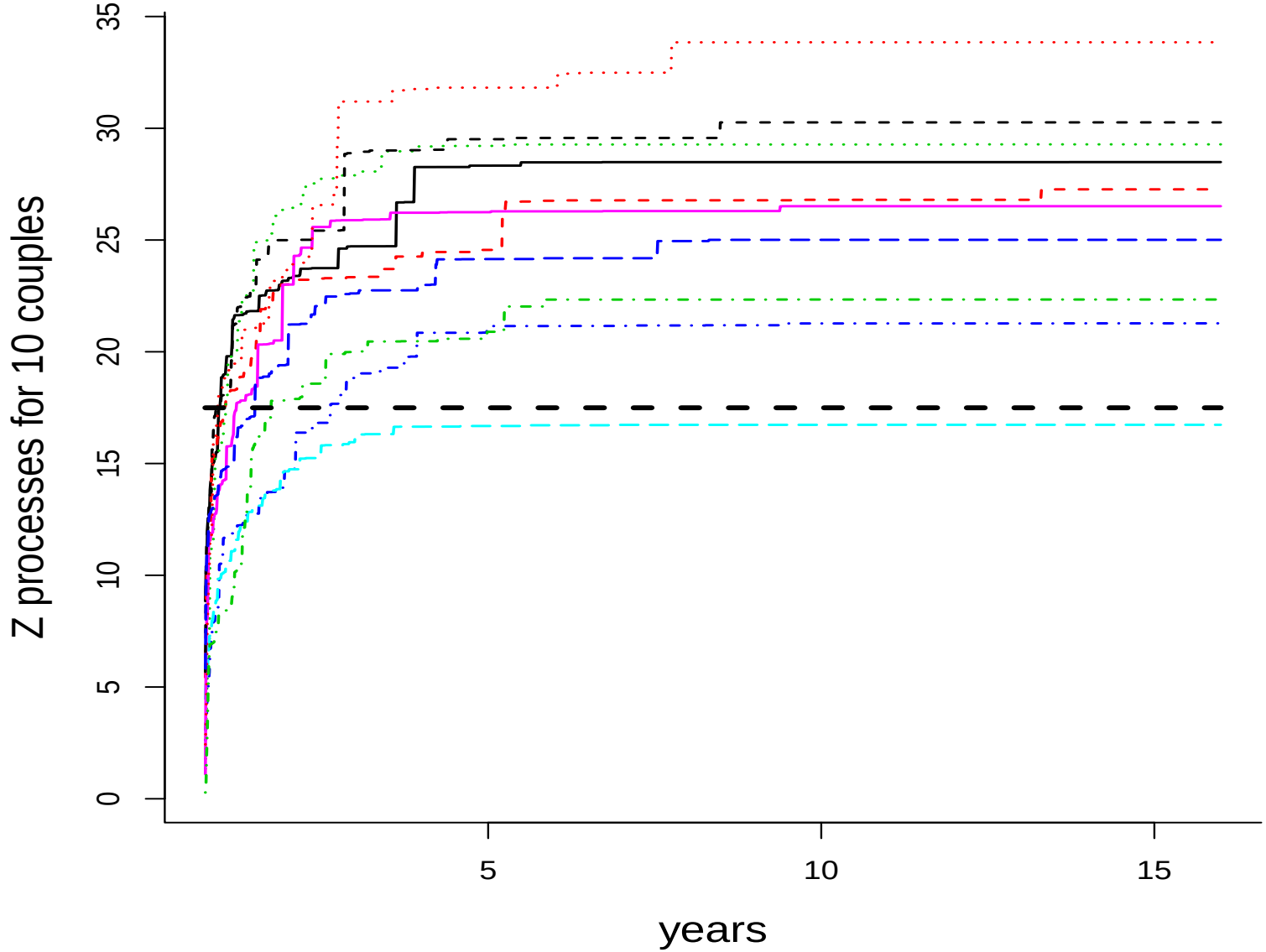
$$h_j = \Pr\{T \in [\ell_j, r_j] \mid T \geq \ell_j\} = 1 - \frac{G(c, aM(r_j), 1)}{G(c, aM(\ell_j), 1)}$$

for the different time intervals.

Model fits very well (via AIC, better than alternatives), also for the $T = \infty$ individuals; cf. cure models.



Empirical and model-fitted hazard rates for the event of a 2nd childbirth, after experiencing a first-born stillbirth, for a population of 451 married Norwegian women.



Simulated Gamma processes for ten couples. The process needs to cross the level $\widehat{c} = 17.45$ (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.

2c: Application C: 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).

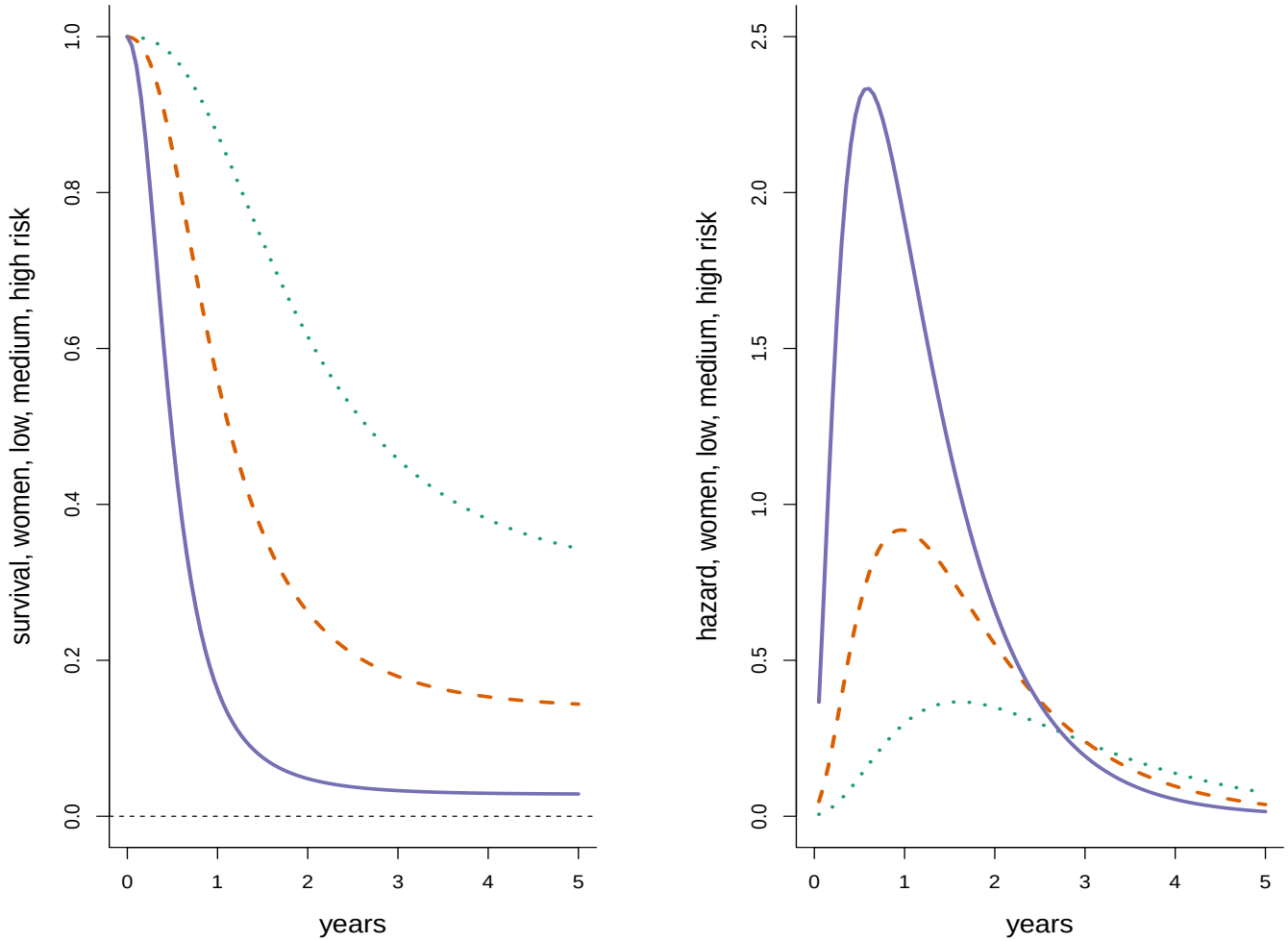
I take the gamma process time-to-hit model

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

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

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

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).

3a: Survival models via Gamma process jumps

Consider a Gamma process, $Z(t) \sim \mathrm{Gam}(A(t), 1)$, with $A(t) = \int_0^t a(s)\,\mathrm{d}s$. There are jumps (mostly small, but some bigger) over each time interval. Suppose an individual is alive as long as all shocks are $\leq v$. Need to find

$$S(t) = \Pr\{T \geq t\} = \Pr\{J(t) < v\},$$

where $J(t)$ is biggest jump over $[0, t]$.

With $Z_m(t) = \sum_{j/m \leq t} G_{m,j}$, and $G_{m,j} \sim \mathrm{Gam}(a(j/m)(1/m), 1)$, we have $Z_m \to_d Z$, and we can prove a Poisson limit:

$$N_m(t) = \sum_{j/m \leq t} I(G_{m,j} > v) \to_d N(t) \sim \mathrm{Pois}(A(t)E_1(v)),$$

with the exponential integral function $E_1(v) = \int_v^\infty (1/u)\exp(-u)\,\mathrm{d}u$. So:

$$S(t) = \Pr\{N(t) = 0\} = \exp\{-A(t)E_1(v)\}.$$

We have 'reinvented' the Cox proportional hazards model, from shocks of a Gamma process – the cumulative hazard rates (can) take the form

$$H_i(t) = A(t)E_1(v_i) = A(t)\exp(x_i^{\mathrm{t}}\beta).$$

Variation I: Suppose individual is alive as long as 3 biggest shocks are below v. Then

$$S_3(t) = \Pr\{N(t) \le 3\} = \exp\{-B(t)\}\{1+B(t)+\tfrac{1}{2}B(t)^2+\tfrac{1}{6}B(t)^3\},$$

with $B(t) = A(t)E_1(v)$. The hazard rate becomes

$$h_3(t) = b(t)\frac{\frac{1}{6}B(t)^3}{1+B(t)+\frac{1}{2}B(t)^2+\frac{1}{6}B(t)^3} = b(t)Q_3(t),$$

with $b(t) = a(t)E_1(v)$ and Q_3 growing from 0 to 1 over time.

Can fit each of $S_1, S_2, S_3, \ldots$ to regression data and determine the mixture proportions, or use AIC or FIC to select the best order.

Variation II: Suppose some individuals 'get used to shocks' (and tolerate more) while others are 'worn out by shocks' (and tolerate less). Assume an individual is alive as long as $G_{m,j} \le v \exp(\gamma w j/m)$, in model formulation above. Then

$$\begin{aligned} S_m(t) &= \prod_{j/m \le t} \Pr\{G_{m,j} \le v \exp(\gamma w j/m)\} \\ &= \prod_{j/m \le t} \{1 - a(j/m)(1/m) E_1(v \exp(\gamma w j/m))\} \\ &\to \exp\Big\{ - \int_0^t a(s) E_1(v \exp(\gamma w s))\, \mathrm{d}s \Big\}. \end{aligned}$$

With survival regression data (t_i, δ_i, x_i), we have an extended Cox model, with hazard rates

$$h_i(t) = a(t) E_1(v_i \exp(\gamma w_i s)), \quad \text{where } E_1(v_i) = \exp(x_i^{\mathrm{t}} \beta),$$

and w_i is one of the covariates. Analysis for given data can provide a confidence curve $\mathrm{cc}(\gamma)$.

3b: Shocks and cumulative shocks, jointly

With a $Z(t) \sim \mathrm{Gam}(A(t), 1)$, suppose an individual is alive as long as $Z(t) < c$ *and* $J(t) < v$, where $J(t)$ is biggest jump experienced over $[0, t]$.

This leads to amenable models if we can derive a formula for the survival, $S(t) = \Pr\{Z(t) < c, J(t) < v\}$. Via $Z_m \to_d Z$, with $Z_m(t) = \sum_{j/m \le t} G_{m,j}$, we have $N_m(t) = \sum_{j/m \le t} I(G_{m,j} > v)$ tending to a Poisson with $A(t)E_1(v)$, and we can prove

$$\begin{aligned} S(t) &= \Pr\{Z(t) < c, N(t) = 0\} \\ &= \int_0^c \Pr\{N(t) = 0 \,|\, z\} g(z, A(t), 1)\, \mathrm{d}z \\ &= \int_0^c \Pr\{J^*(t) < zv\} g(z, A(t), 1)\, \mathrm{d}z, \end{aligned}$$

where $J^*(t)$ is the biggest jump in a certain Dirichlet process $D^*(\cdot)$ over $[0, t]$. Can be done, via Hjort and Ongaro (2006) $\Longrightarrow$ full inference.

3c: Life is full of dangers

Suppose an invidual lives a life full of independent competing dangers, with cause j of event stemming from one of

$$Z_j(t) \sim \mathrm{Gam}((1/m)b(j/m)M(t), 1) \text{ crossing threshold } c(j/m).$$

Then with $T = \min(T_1, \ldots, T_m)$, and m big,

$$\begin{aligned} S(t) &= \Pr\{\text{each } Z_j(t) < c\Big(\frac{j}{m}\Big)\} = \prod_{j \le m} G\Big(c\Big(\frac{j}{m}\Big), \frac{1}{m} b\Big(\frac{j}{m}\Big) M(t), 1\Big) \\ &= \prod_{j \le m} \Big\{1 - \frac{1}{m} b\Big(\frac{j}{m}\Big) M(t) E_1\Big(c\Big(\frac{j}{m}\Big)\Big) + O(1/m^2)\Big\} \\ &\to \exp\Big\{-M(t) \int_0^1 b(s) E_1(c(s))\, \mathrm{d}s\Big\}. \end{aligned}$$

This leads to a large class of plausible models, where special subclasses may be used for a set of given data.

4: Extended logistic regression

Standard logistic regression:

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

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

Extension:

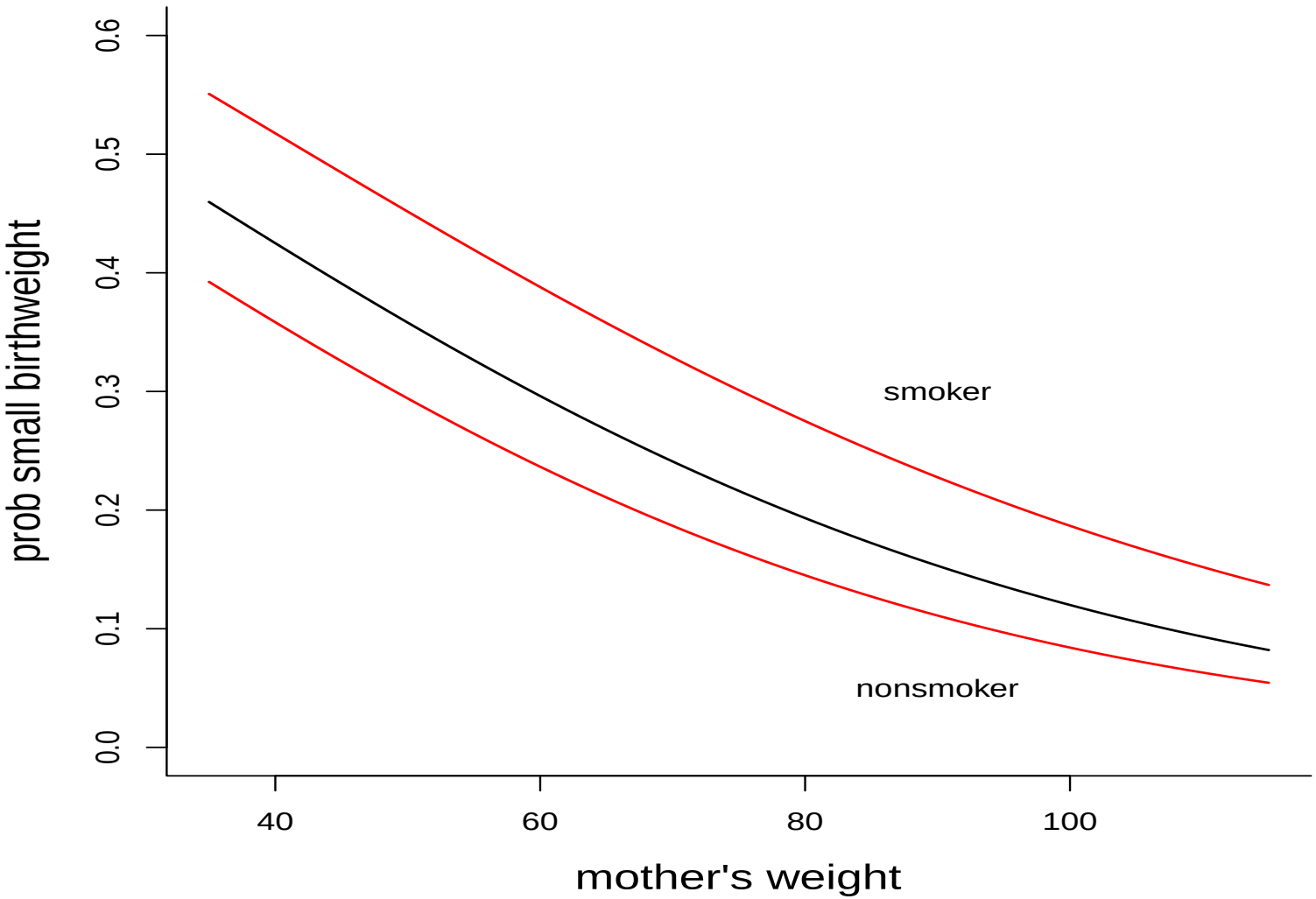
$$p_i = \Pr(Y_i = 1 \,|\, x_i, z_i) = G\big(\log\{1+\exp(x_i^{\mathrm{t}}\beta)\}, a_i, 1\big),$$

where $a_i = a(z_i)$. Could have $a_i = \exp(z_i^{\mathrm{t}}\gamma)$, 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,

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



5a: 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

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

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

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

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

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

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

5b: War of Roses (1455-1487) and Game of Thrones

We have 400 noblemen from the two universes. They die of violence or of natural causes.

- WoR: 126 dead men, 35% violence
- GoT: 274 dead men, 56 alive, 81% violence

We use two competing risk Gamma processes and time to hit:

$$Z_n(t) \sim \text{Gam}(a_n t^{\kappa_n}, 1) \quad \text{and} \quad Z_v(t) \sim \text{Gam}(a_v t^{\kappa_v}, 1).$$

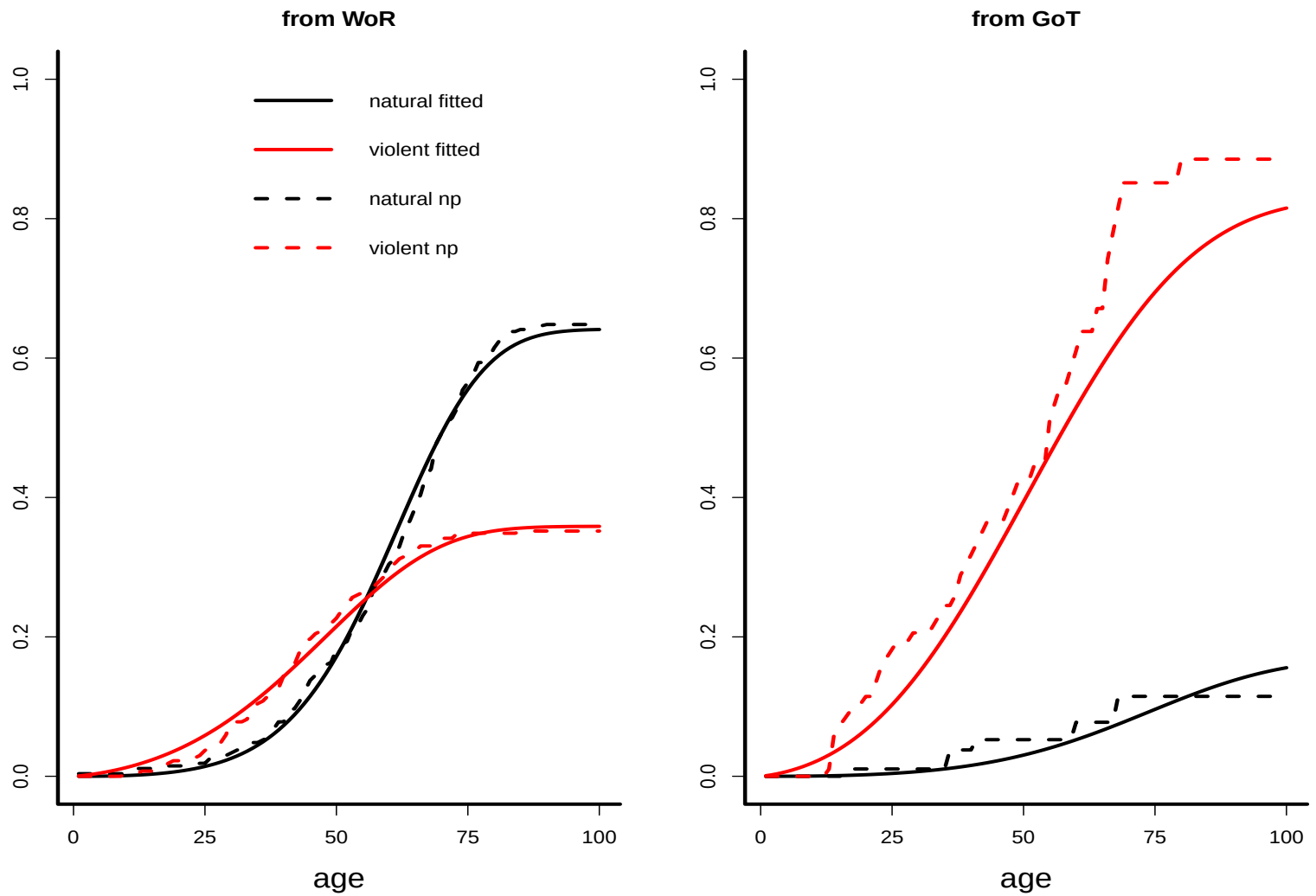
Our model uses also L, the length of the wikipedia article:

$$\alpha_n = \exp\{\beta_{n,0} + \beta_{n,1} I(GoT)\},$$
$$\alpha_v = \exp\{\beta_{v,0} + \beta_{v,1} I(GoT) + \beta_{v,2} L + \beta_{v,3} I(GoT) \times L\}.$$

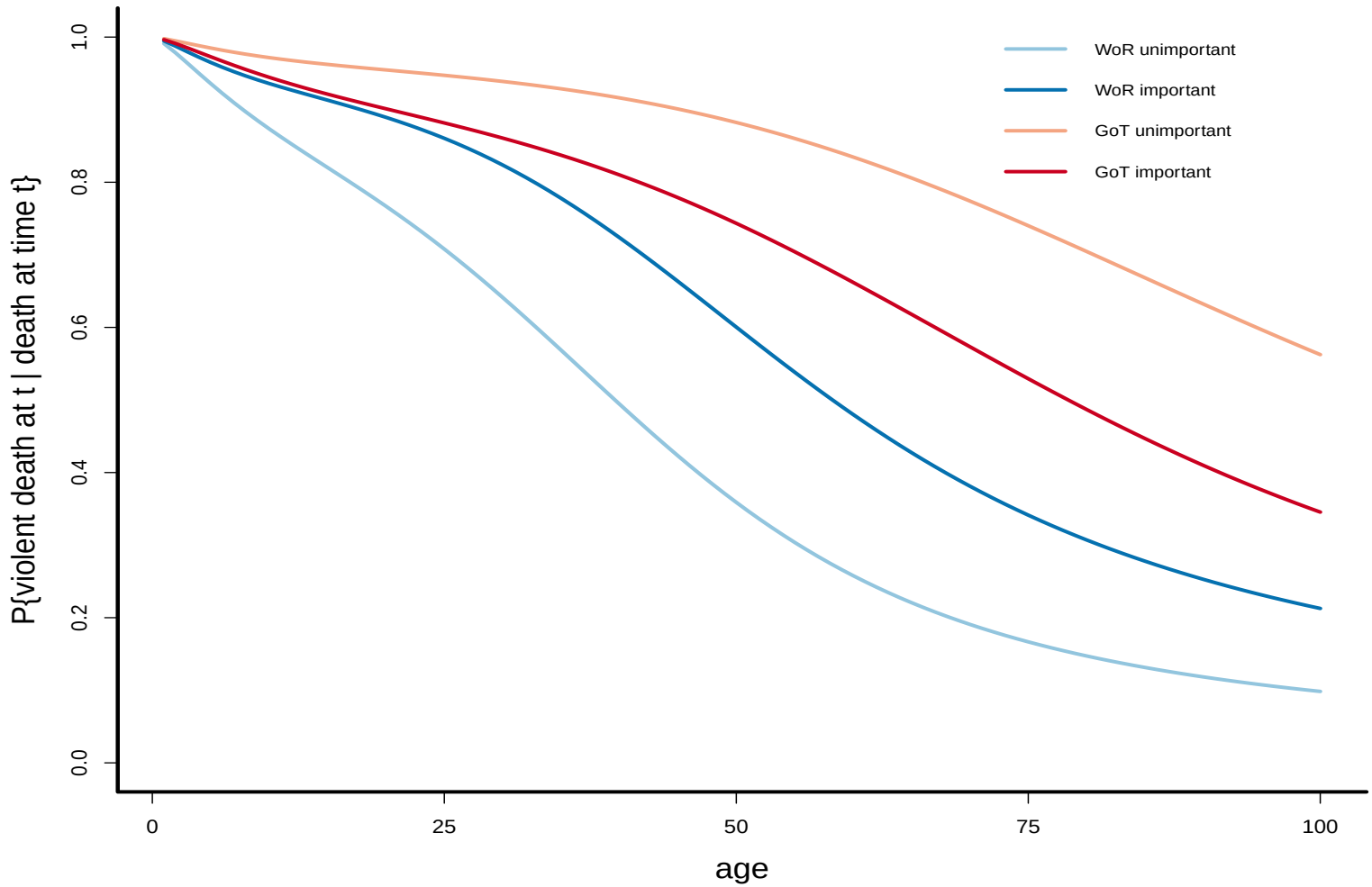
Inference based on log-likelihood:

$$\ell(\theta) = \sum_{\delta_i=0} \log S(t_i \,|\, \theta) + \sum_{\delta_i=1} \log f_1^*(t_i \,|\, \theta) + \sum_{\delta_i=2} \log f_2^*(t_i \,|\, \theta).$$

Fitted cumulative incidence functions compared with nonparametric estimators:



The probability of dying of cause j given death at time t: $q_j(t) = f_j^*(t)/\{f_1^*(t) + \cdots + f_k^*(t)\}$. Here: $\Pr\{\text{dies a violent death at } t \,|\, \text{dies at } t\}$, in two universes.



6: Frailtifying the Gamma process threshold model

My favourite Gamma process threshold model is: event takes place when $Z(t) \geq c$, where $Z(t) \sim \mathrm{Gam}(aM_0(t), 1)$:

$$S(t \mid c) = \Pr\{T \geq t \mid c\} = G(c, aM_0(t), 1).$$

Frailty: give c a distribution with distribution $F_0(c) = 1 - S_0(c)$. Then, observed in the population:

$$\begin{aligned} S(t) &= \int_0^\infty S(t \mid c)\, \mathrm{d}F_0(c) \\ &= \int \int I(x \leq c) g(x, aM_0(t), 1)\, \mathrm{d}x\, \mathrm{d}F_0(c) \\ &= \int S_0(x) \frac{1}{\Gamma(aM_0(t))} x^{aM_0(t)-1} \exp(-x)\, \mathrm{d}x. \end{aligned}$$

Frailty for thresholds translates to downweighting over time of the gamma density. (Can also frailtify over a.)

Special case: $c \sim \mathrm{Expo}(\gamma)$ implies $S(t) = \exp\{-bM_0(t)\}$, with $b = a \log(1 + \gamma)$.

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, Lee, Whitmore, 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.

7. Two-stage models for events: (i) first $Z_1(t) \sim \mathrm{Gam}(M(t), 1)$ at work, until threshold c_1 at time T_1; (ii) then $Z_2(t) \sim \mathrm{Gam}(M^*(t), 1)$ sets in, with different M^*, and might hit c_2. Links to cure models.

8. Gamma-Gamma process, to reflect more uncertainty (or random effects): $Z \,|\, Z_0 \sim \mathrm{Gam}(Z_0, 1)$ and $Z_0 \sim \mathrm{Gam}(M, 1)$. Then

$$\mathrm{E}\, Z(t) = M(t) \quad \text{and} \quad \mathrm{Var}\, Z(t) = M(t) + M(t).$$

Can again work with time-to-threshold and time-to-jumpsize.

9. Excessive risk in some time periods:

$$\mathrm{d}Z_i(t) = \mathrm{d}Z_0(t) + x_i(t)\,\mathrm{d}R(t) = \begin{cases} \mathrm{d}Z_0(t) & \text{when normal,} \\ \mathrm{d}Z_0(t) + \mathrm{d}R(t) & \text{when danger.} \end{cases}$$

With Gamma processes for Z_0 and R, can make inference for their parameters.

(Some) references

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.

Cunen, C. and Hjort, N.L. (2018). Survival and event history models via gamma processes.

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.

Hjort, N.L. and Ongaro, A. (2006). On the distribution of random Dirichlet jumps. *Metron*.

Lee, M.-L.T. and Whitmore, G.A. (2006). Threshold regression for survival analysis: *Satistical Science.*

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