Survival Densities & Regression from Phase-Type Models

Eric V. Slud, Census Bureau & Univ. of Maryland
Jiraphan Suntornchost, Univ. of Maryland

OUTLINE

I. History : Survival densities & Phase-type Models

II. Motivations for Parametric Densities

III. Phase-type representations & parameterizations

IV. Model-fitting, Software & References

V. Data analysis: SEER Breast Cancer Data 1992-2002

Actuarial Sources

Features of hazard curves: $\begin{cases} q_x = & \textit{age-specific death rate} \\ \mu_x = & \textit{force of mortality} = \textit{hazard} \end{cases}$

- increasing, decreasing, or 'bathtub' shapes
- Gompertz (1825) & Makeham (1864): $q_x = A + Bc^x$
- power law, Weibull (1939): $\mu_x = \beta \, \alpha \, x^{\alpha - 1}$

These models unified by:

Fréchet-Fisher-Tippett-Gnedenko (1927-1948) Theorem
characterizing distribution limits of $\max_{1 \leq i \leq n} X_i$ for iid r.v.'s

But practical actuaries rely on 'Graduation'
(Whittaker-Henderson smoothing splines) to fit q_x

What we would call a nonparametric approach ...

Demography

Heiligman-Pollard (1980) eight-parameter model:

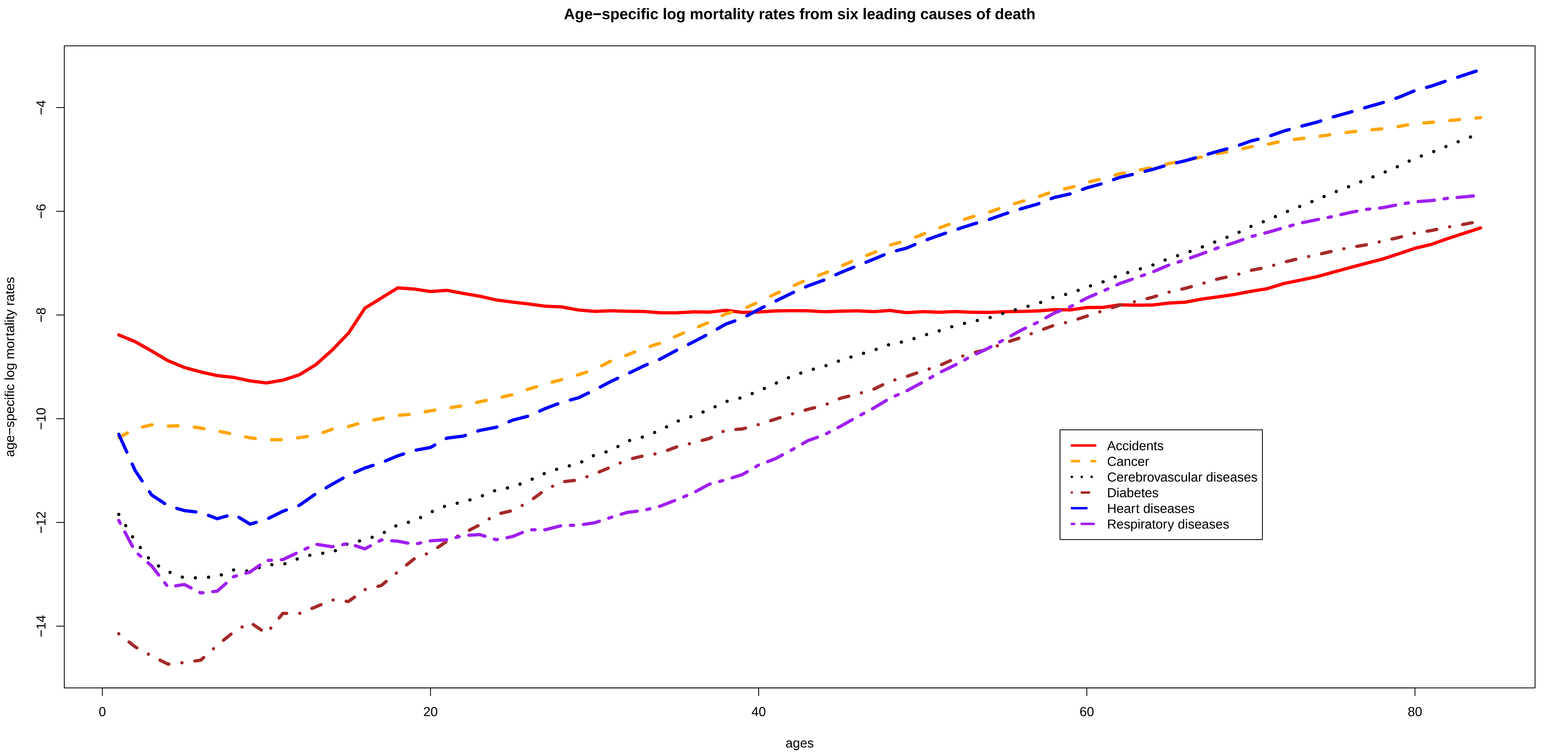
$$\frac{q_x}{1-q_x} = A^{(x+B)^C} + D\,\exp(-E\log^2\frac{x}{F}) + GH^x$$

parameters not interpretable, messy to fit, not really practical

Demographers also use a highly parameterized, effectively nonparametric model:

Lee-Carter (1992): $\log q_{x,t} = \alpha_x + \beta_x\,\gamma_t + \epsilon_{x,t}$

widely used as benchmark model, generalizes linear t-projections of x-mortality (often $\gamma_t \approx$ linear)



Age−specific log mortality rates from six leading causes of death
age−specific log mortality rates
−4
−6
−8
−10
−12
−14
0
20
40
60
80
ages
Accidents
Cancer
Cerebrovascular diseases
Diabetes
Heart diseases
Respiratory diseases

Digression: Extension of Lee-Carter Cause-Specific & Forecasted Curves

Suntornchost, Slud & Wei (2011) find intervals of ages x with **different** time-profiles γ_t for distinct death-causes.

Fit spline-smoothed γ_t in each age-group
to NCHS cause-specific mortality curves.

Use disaggregated, not combined, model to forecast in t.

Parameterization reduced when data are cross-classified.

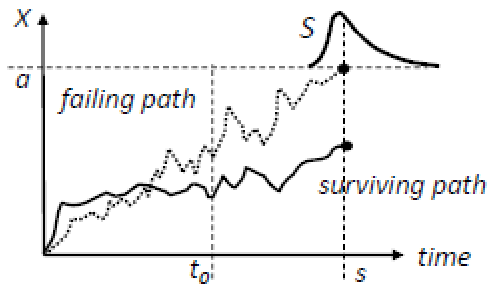
Threshold-crossing (Cum. Damage) Models

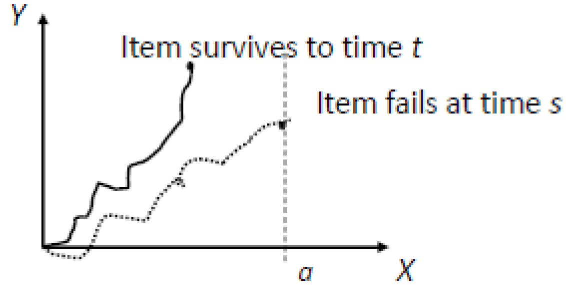
Models in **reliability** (Singpurwalla 1996): $X(t)$ underlying unobservable stochastic process (‘degradation’, ’damage’)

Failure occurs when $X(t)$ crosses threshold a
Correlated process $Y(t)$ may be observed

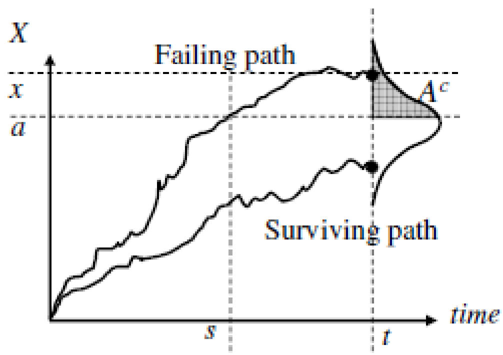
If $X(t)$ is Wiener process with drift, waiting-time T to cross is 2-parameter **Inverse Gaussian**: extended reflection principle gives $P(T \leq t) = P(\max_{0 \leq s \leq t} X(s) \geq a)$

General approach to survival hazards: Aalen & Gjessing (2001), Regression models: Lee & Whitmore (Stat Sci 2006)





Pictures of paths for Degradation Process X in relation to Failure



Slide from Vasilis Sotiris thesis presentation on simultaneous regression models for degradation & failure.

Biomedical Models $\rightarrow$ Markov Chains

Healthy $\mapsto$ Illness $\mapsto$ Death progressions (Infectious Disease)

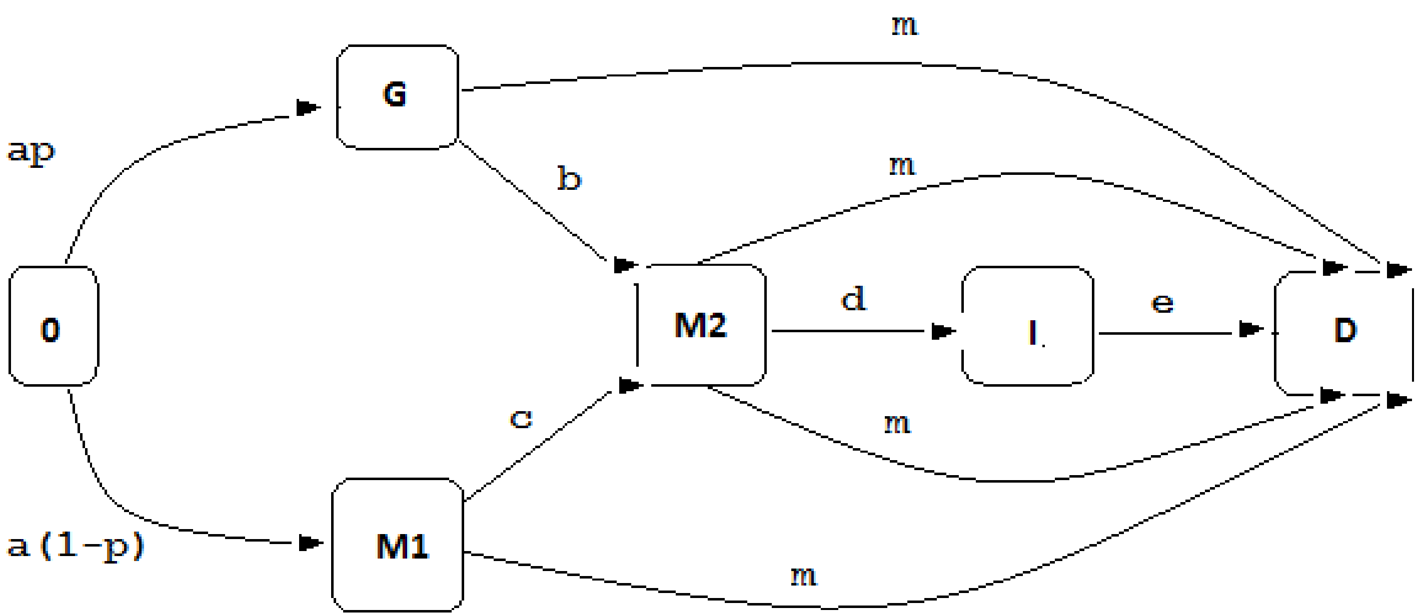
Latent State, Multihit models: internal transformations of cells (developmental disease, Cancer)

Armitage & Doll (1954) observed power law for cancer incidence: $P(T_I \leq t) \approx c\, t^k$ over range of t, suggested T_I distributed as sum of indep. Expon. waiting times

$$\boxed{\mathbf{0}} \xrightarrow{\lambda_1} \boxed{1} \xrightarrow{\lambda_2} \cdots \xrightarrow{\lambda_{k-1}} \boxed{\text{k-1}} \xrightarrow{\lambda_k} \boxed{\text{k}}$$

Moolgavkar (2004): multistage cancer causation models now explanatory, supported by genetic/biologic evidence

Knudsen 1954 Retinoblastoma model



Phase-Type Models

<u>Definition</u>: a **phase-type** r.v. is the absorption time into death-state in continuous-time homogeneous Markov chain.

Initial state is **0**, terminal state **D**, other states $\{1, \ldots, k\}$, $(k+2) \times (k+2)$ transition intensities $Q = \{q_{ab}\}$, $\quad Q\mathbf{1} = \mathbf{0}$

$$P(T \le t) \;=\; (\exp(tQ))_{0D} \;=\; P_{0D}(t)$$

Origins in applied probability, Queueing (M. Neuts 1981) and Compartmental Models in pharmacokinetics.

Phase-type Representations

Continuous-time constant hazard state transitions represent many statistical modeling constructions: class is closed under

sums of indep. r.v.'s, mixtures, min's and max's of indep.'s

Inverse-Gaussian and other diffusion boundary-crossing times obtained as approximate absorption-times:

random walk with drift: up-steps rate $p\lambda$**, down** $(1-p)\lambda$

State (j, m) denotes m net up-steps after $j \geq 0$ transitions. Define all states with $m \geq A$ to be lumped as **D** death-state.

Process approximates Wiener-process trajectory with time $t = j/\lambda$, $\sigma^2 = p(1-p)$, drift $(2p-1)\lambda$

Why Parametric Densities ?

Even though nonparametric methods in biostatistics (semiparametric regression models) are available,

parametric survival models still have useful role when:

- subjects are highly cross-classified with widely varying prognosis, as in cancer databases like SEER, or
- different covariates might influence different steps in multistage illness/death pathways, or
- researchers are looking for hints whether different phenomena are operating in subpopulations (mixtures).

Parsimony may require many of the phase-type transition rates to be common or related.

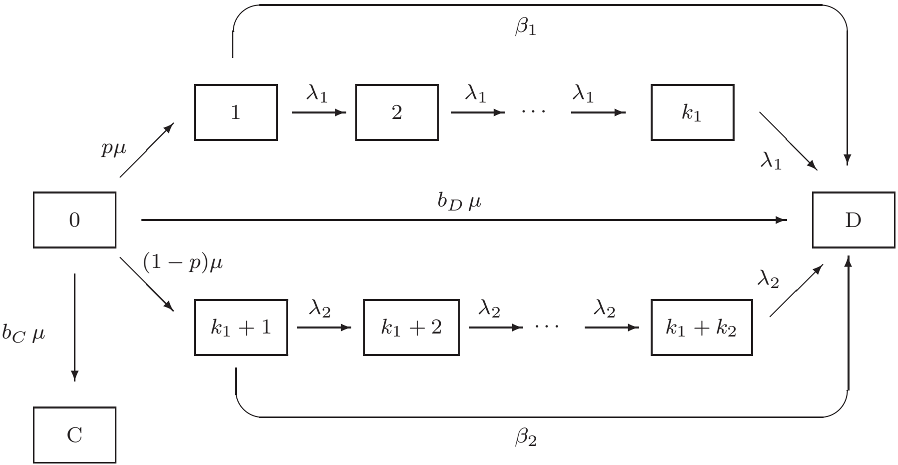


Fig. 1 Markov transition diagram for `Model F` with immediate cures and failures, additional direct failures from states 1, 2, and two failure pathways.

Example, Special Phase-Type Model

The Phase-Model Picture just shown has the features:

After waiting for time $T_1 \sim \text{Expon}((1 + b_C + b_D)\mu)$,

$$O \mapsto C,\, D, 1, k+1 \quad \text{with prob.'s} \quad \frac{(b_C,\, b_D,\, p,\, 1-p)}{1 + b_C + b_D}$$

From state 1, absorption time to D is a mixture
w. prob. $q_1 = \beta_1/(\beta_1 + \lambda_1)$, $\quad = T_{1D} \sim \text{Expon}(\lambda_1 + \beta_1)$,
w. prob. $1 - q_1$, $\quad = T_{1D} + G_1 \;, G_1 \sim \text{Gamma}(k_1 - 1, \lambda_1)$

Similarly, cond'l absorption time from state 2 is a mixture
w. prob. q_2, $1 - q_2$ of $T_{2D} \sim \text{Expon}(\lambda_2 + \beta_2)$ and $T_{2D} + G_2$,
where $q_2 = \beta_2/(\beta_2 + \lambda_2)$, $\quad G_2 \sim \text{Gamma}(k_2 - 1, \lambda_2)$.

Computing Formulas for Likelihood in Model F

$$P_{OD}(t) \;=\; \frac{b_D}{1+b_C+b_D}\,(1-e^{-\mu(1+b_C+b_D)t})$$

$$+\;\sum_{j=1}^{2} \frac{p^{2-j}(1-p)^{j-1}}{1+b_C+b_D}\,\Big[q_j\,\mathsf{Exp}(\mu(1+b_C+b_D)) * \mathsf{Exp}(\beta_j+\lambda_j)(t)$$

$$+\;(1-q_j)\,\mathsf{Exp}(\mu(1+b_C+b_D)) * \mathsf{Exp}(\beta_j+\lambda_j) * \mathsf{Gam}(k_j,\lambda_j)(t)\Big]$$

where for $S \sim \mathsf{Exp}(a)$, $T \sim \mathsf{Exp}(b)$, $U \sim \mathsf{Gam}(r,\lambda)$,

$$f_{S+T}(t) \;=\; \frac{ab}{b-a}\,(e^{-at} \;-\; e^{-bt}) \;\;, \qquad\qquad f_{S+U}(t) \quad \text{also explicit}$$

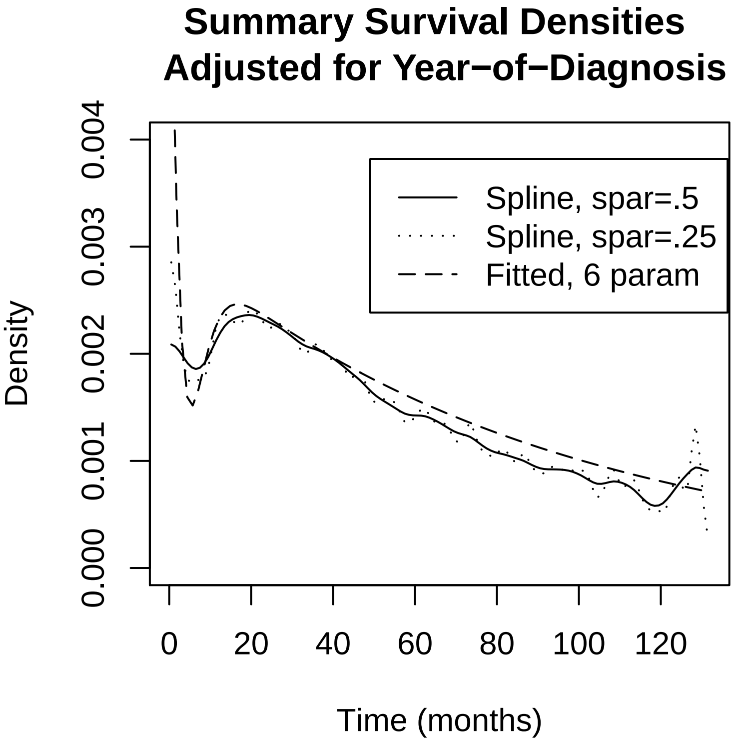
Breast Cancer Data Analysis Using Model F

Data from SEER cancer database 1992-2002,
as in Anderson et al. (2006): motivation of that paper is separating post-diagnosis mortality by **Estrogen-Receptor** (ER) status

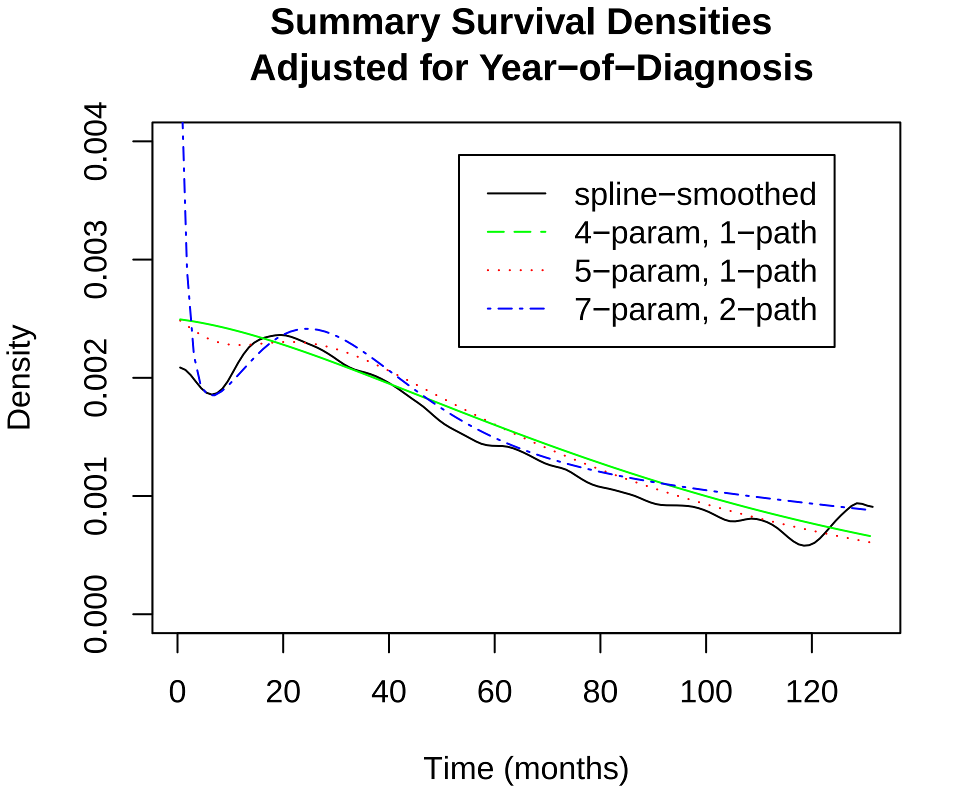
analyzed data on 198,785 white female breast-cancer cases *from time of diagnosis*

• first fit proportional hazards (Cox) model to remove effect of *Diagnosis Year*

• then produced summary survival curve, fitted smoothing spline, produced density.



Summary Survival Densities
Adjusted for Year−of−Diagnosis
Spline, spar=.5
Spline, spar=.25
Fitted, 6 param
Density
0.000
0.001
0.002
0.003
0.004
0
20
40
60
80
100
120
Time (months)


Summary Survival Densities
Adjusted for Year−of−Diagnosis
Density
0.000
0.001
0.002
0.003
0.004
0
20
40
60
80
100
120
Time (months)
spline−smoothed
4−param, 1−path
5−param, 1−path
7−param, 2−path

EM algorithm (Asmussen et al. 1996)

Consider the embedded Markov chain $I_0, I_1, \ldots, I_{M-1}$ $(I_M = D)$, and sojourn times $S_0, S_1, ..., S_{M-1}$.

$\mathbf{y} = (y_1, y_2, ..., y_N)$, sample of phase-type observed times

'Complete observation': $\mathbf{x} = (i_0, i_1, \ldots, i_{M-1}, s_0, \ldots, s_{M-1})$ and sojourn times satisfy $y = s_0 + s_1 + s_2 + ... + s_{M-1}$.

Transient states are $\{1, \ldots, p\}$, absorbing D.

Transition intensity matrix is $\mathbf{Q} = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix}$

$$\mathbf{T}_{jk} \equiv t_{jk} = q_{jk}, \;\; 1 \leq j, k \leq p, \qquad t_j \equiv q_{jD}$$

$$p_{jk} = \mathbf{P}(I_{n+1} = k \,|\, I_n = j) \; = \; t_{jk}/(-q_{jj}) \; , \quad k \in \{1, ..., p, D\} \setminus \{j\}$$

The density of complete sample $(\mathbf{x})$ can be written

$$f(\mathbf{x};\pi,\mathbf{T}) = \prod_{n=1}^{N}\Big(\prod_{i=1}^{p}\pi_i^{B_i^{(n)}}\prod_{i=1}^{p}\exp(t_{ii}^{(n)}Z_i^{(n)})\prod_{i=1}^{p}\prod_{j=0,j\neq i}^{p}(t_{ij}^{(n)})^{N_{ij}^{(n)}}\Big),$$

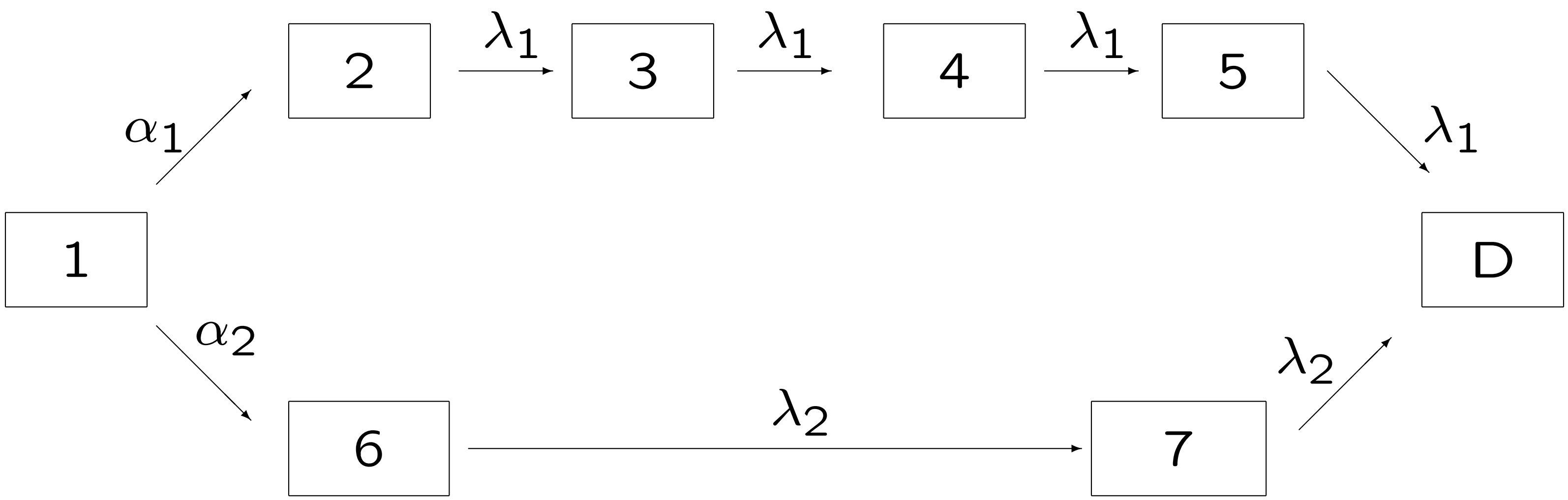
where $\pi =$ initial distribution for the Markov Chain, and

$$B_i^{(n)} = I_{\{I_0^{(n)}=i\}}$$

$$Z_i^{(n)} = \sum_{k=0}^{m(n)-1} I_{\{I_k^{(n)}=i\}}S_k^{(n)} = \text{ total time the process spends in state } i$$

$$N_{ij}^{(n)} = \sum_{k=0}^{m(n)-1} I_{\{I_k^{(n)}=i,I_{k+1}^{(n)}=j\}} = \text{number of jumps from state } i \text{ to } j$$

Special case



The mixture of $\mathrm{Exp}(\alpha_1) * \mathrm{Gamma}(4, \lambda_1)$ and $\mathrm{Exp}(\alpha_2) * \mathrm{Gamma}(2, \lambda_2)$.

The log-likelihood function is given by

$$
\begin{aligned}
l = \log(\alpha_1) \sum_{n=1}^{N} N_{12}^{(n)} &+ \log(\lambda_1) \sum_{n=1}^{N} (\sum_{i=2}^{4} N_{i,(i+1)}^{(n)} + N_{5D}^{(n)}) \\
&+ \log(\alpha_2) \sum_{n=1}^{N} N_{16}^{(n)} + \log(\lambda_1) \sum_{n=1}^{N} (N_{67}^{(n)} + N_{7D}^{(n)}) \\
&- (\alpha_1 + \alpha_2) \sum_{n=1}^{N} Z_1^{(n)} - \lambda_1 \sum_{n=1}^{N} (\sum_{i=2}^{5} Z_i^{(n)}) \\
&- \lambda_2 \sum_{n=1}^{N} (Z_6^{(n)} + Z_7^{(n)}).
\end{aligned}
$$

E-step

Unknown 'parameters' $N_{ij}^{(n)}$, $Z_i^{(n)}$ for uncensored observations $(1 \le n \le N)$ are replaced by conditional expectations given observed data as in Asmussen (1996) :

$$E_{(\pi,\mathbf{T})}\big(Z_i^{(n)}|Y = y_n\big) = c_i(y_n : i|\pi,\mathbf{T})/(\pi\,\mathbf{b}(y_n|\mathbf{T}))$$
$$E_{(\pi,\mathbf{T})}\big(N_{ij}^{(n)}|Y = y_n\big) = t_{ij}^{(n)}\,c_j(y_n : i|\pi,\mathbf{T})/(\pi\,\mathbf{b}(y_n|\mathbf{T}))$$
$$E_{(\pi,\mathbf{T})}\big(N_{iD}^{(n)}|Y = y_n\big) = t_i\,a_i(y_n|\pi,\mathbf{T})/(\pi\,\mathbf{b}(y_n|\pi,\mathbf{T})),$$

where $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}(\cdot\,; i|\cdot)$ for $1 \le i \le p$ are p-dim vector functions

$$\mathbf{a}(y|\pi,\mathbf{T}) \;=\; \pi\ \text{exp}(\mathbf{T}y) \quad , \qquad \mathbf{b}(y|\pi,\mathbf{T}) \;=\; \text{exp}(\mathbf{T}y)\,\mathbf{t}$$

$$\mathbf{c}(y; i|\pi,\mathbf{T}) \;=\; \int_0^y \Big\{\pi\ \text{exp}(\mathbf{T}u)\,\mathbf{e}_i\Big\}\,\Big\{\text{exp}(\mathbf{T}(y-u))\,\mathbf{t}\Big\}\,du$$

The unknowns $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ are obtained by solving ordinary differential equations, by the Runge-Kutta numerical method :

$$\begin{aligned}
\mathbf{a}'(y|\pi,\mathbf{T}) &= \mathbf{a}(y,\pi,\mathbf{T})\,\mathbf{T} \\
\mathbf{b}'(y|\pi,\mathbf{T}) &= \mathbf{T}\,\mathbf{b}(y,\pi,\mathbf{T}) \\
\mathbf{c}'(y,i|\pi,\mathbf{T}) &= \mathbf{T}\,\mathbf{c}(y,i|\pi,\mathbf{T}) + a_i(y,\pi,\mathbf{T})\,\mathbf{T} \ , \qquad i = 1,..,p
\end{aligned}$$

M-Step

The likelihood is maximized, and ML estimates are given as:

$$\widehat{\alpha}_1 = \frac{\sum_{n=1}^{N} N_{12}^{(n)}}{\sum_{n=1}^{N} Z_1^{(n)}} \quad , \quad \widehat{\lambda}_1 = \frac{\sum_{n=1}^{N} (\sum_{i=2}^{4} N_{i,(i+1)}^{(n)} + N_{5D}^{(n)})}{\sum_{n=1}^{N} (\sum_{i=2}^{5} Z_i^{(n)})}$$

$$\widehat{\alpha}_2 = \frac{\sum_{n=1}^{N} N_{16}^{(n)}}{\sum_{n=1}^{N} Z_1^{(n)}} \quad , \quad \widehat{\lambda}_2 = \frac{\sum_{n=1}^{N} (N_{67}^{(n)} + N_{7D}^{(n)})}{\sum_{n=1}^{N} (Z_6^{(n)} + Z_7^{(n)})}.$$

Fisher Information by an EM algorithm

By Oakes (1999), the Fisher information matrix is given by substituting $\theta_1 = \widehat{\theta}$ into

$$\frac{\partial^2 l(\theta_1 \; ; y)}{\partial \theta_1^2} = \left\{ \frac{\partial^2 Q(\theta_2|\theta_1)}{\partial \theta_2^2} + \frac{\partial^2 Q(\theta_2|\theta_1)}{\partial \theta_1 \partial \theta_2} \right\}|_{\theta_2=\theta_1},$$

where

$$Q(\theta_2|\theta_1) = \mathbb{E}_{\theta_1}(l(\theta_2\,; \mathbf{z}) \mid \mathbf{y}),$$

and $z = (z_1, ..., z_N)$ denotes the complete dataset for the N observations.

$$
\begin{aligned}
Q(\widehat{\theta} \mid \theta_1) &= \mathbb{E}_{\theta_1}(l(\theta_2; \mathbf{z}) \mid \mathbf{y}) \Big|_{\theta_2=\widehat{\theta}} \\
&= \log(\widehat{\alpha}_1) \sum_{n=1}^{N} \mathbb{E}(N_{12}^{(n)}|\mathbf{y}) + \log(\widehat{\lambda}_1) \sum_{n=1}^{N} \mathbb{E}(\sum_{i=2}^{4} N_{i,(i+1)}^{(n)} + N_{5D}^{(n)}|\mathbf{y}) \\
&\quad + \log(\widehat{\alpha}_2) \sum_{n=1}^{N} \mathbb{E}(N_{16}^{(n)}|\mathbf{y}) + \log(\widehat{\lambda}_2) \sum_{n=1}^{N} \mathbb{E}(N_{67}^{(n)} + N_{7D}^{(n)}|\mathbf{y}) \\
&\quad - (\widehat{\alpha}_1 + \widehat{\alpha}_2) \sum_{n=1}^{N} \mathbb{E}(Z_1^{(n)}|\mathbf{y}) - \widehat{\lambda}_1 \sum_{n=1}^{N} \mathbb{E}(\sum_{i=2}^{5} Z_i^{(n)}|\mathbf{y}) \\
&\quad - \widehat{\lambda}_2 \sum_{n=1}^{N} \mathbb{E}(Z_6^{(n)} + Z_7^{(n)}|\mathbf{y})
\end{aligned}
$$

With the notations:

$$\mathbf{ME}^{(n)} = \begin{pmatrix} c_1(y_n : 1|\pi, \mathbf{T}) & c_2(y_n : 1|\pi, \mathbf{T}) \\ c_{25}(y_n|\pi, \mathbf{T}) & c_{52}(y_n|\pi, \mathbf{T}) \\ c_1(y_n : 1|\pi, \mathbf{T}) & c_6(y_n : 1|\pi, \mathbf{T}) \\ c_{67}(y_n|\pi, \mathbf{T}) & c_{76}(y_n|\pi, \mathbf{T}) \end{pmatrix},$$

where

$$c_{25}(y_n|\pi, \mathbf{T}) := \sum_{i=2}^{5} c_i(y_n : i|\pi, \mathbf{T})$$

$$c_{52}(y_n|\pi, \mathbf{T}) := \sum_{i=2}^{4} c_{i+1}(y_n : i|\pi, \mathbf{T}) + a_5(y_l : \pi, \mathbf{T})$$

$$c_{67}(y_n|\pi, \mathbf{T}) := \sum_{i=6}^{7} c_i(y_n : i|\pi, \mathbf{T})$$

$$c_{76}(y_n|\pi, \mathbf{T}) := c_7(y_n : 6|\pi, \mathbf{T}) + a_7(y_l : \pi, \mathbf{T})$$

The Fisher Information matrix

$$\mathbf{L} = (L_{ij}) = -\Big(\frac{\partial^2 Q(\phi\,|\,\theta)}{\partial\phi_j\partial\phi_i} + \frac{\partial^2 Q(\phi\,|\,\theta)}{\partial\theta_j\partial\phi_i}\Big)_{\phi=\theta} \qquad \text{is given by}$$

$$\begin{aligned} L_{ij} = &\sum_{n=1}^{N} \frac{1}{f(y_n)}\Big[\frac{\partial}{\partial\theta_j}\mathsf{ME}_{i2}^{(n)} - \frac{\partial}{\partial\theta_j}\mathsf{ME}_{i1}^{(n)}\Big] \\ &- \sum_{n=1}^{N} \frac{1}{(f(y_n))^2}\frac{\partial}{\partial\theta_j}f(y_n)\Big[\mathsf{ME}_{i2}^{(n)} - \mathsf{ME}_{i1}^{(n)}\Big], \end{aligned}$$

To find the Fisher Information matrix numerically, we solve the following two systems of equations:

$$\frac{d}{dy}\mathbf{e}(y|\pi,\mathbf{T}) = \mathbf{T}\mathbf{e}(y|\pi,\mathbf{T})$$

$$\frac{d}{dy}\mathbf{T}_\theta^{(n)}(y) = \mathbf{e}(y|\pi,\mathbf{T})\frac{\partial \mathbf{T}}{\partial\theta} + \mathbf{T}_\theta^{(n)}(y)\mathbf{T}$$

$$\frac{d}{dy}\mathbf{C}^{(n)}(y|\pi,\mathbf{T}) = \mathbf{T}\mathbf{C}^{(n)}(y|\pi,\mathbf{T}) + \mathbf{t}\otimes(\pi\mathbf{e}(y|\pi,\mathbf{T}))$$

$$\begin{aligned}\frac{d}{dy}\mathbf{C}_\theta^{(n)}(y|\pi,\mathbf{T}) &= \Big(\mathbf{t}\otimes\big(\pi\mathbf{T}_\theta^{(n)}(y)\big) + (\frac{\partial\mathbf{t}}{\partial\theta})\otimes\big(\pi\mathbf{e}(y|\pi,\mathbf{T})\big)\Big)\\ &\quad + \mathbf{T}\mathbf{C}_\theta^{(n)}(y|\pi,\mathbf{T}) + \frac{\partial\mathbf{T}}{\partial\theta}\mathbf{C}^{(n)}(y|\pi,\mathbf{T})\end{aligned}$$

where $\mathbf{e}(y|\pi,\mathbf{T}) = \exp(\mathbf{T}y)$, $\mathbf{T}_\theta^{(n)}(y_n) = \dfrac{\partial}{\partial\theta}\exp(\mathbf{T}y_n)$, and $\otimes$ denotes a Kronecker product. The system of differential equations can be solved by the Runge Kutta method with the initial value $\mathbf{e}(0|\pi,\mathbf{T}) = \mathbf{I}_p$, and $\mathbf{C}^{(n)}(0|\pi,\mathbf{T}) = \mathbf{C}_{\alpha_i,2}^{(n)}(0|\pi,\mathbf{T}) = \mathbf{C}_\theta^{(n)}(0|\pi,\mathbf{T}) = \mathbf{T}_\theta^{(n)}(0) = \mathbf{O}_p$ for all $\theta \in \{\alpha_1,\lambda_1,\alpha_2,\lambda_2\}$.

Numerical Results

We consider the mixture of $\text{Exp}(\alpha_1) * \text{Gamma}(4, \lambda_1)$ and $\text{Exp}(\alpha_2) * \text{Gamma}(2, \lambda_2)$, with sample size of 100.

Parameters	True values	MLE	SD
α_1	0.15	0.06477	0.05564
λ_1	0.15	0.91316	0.50344
α_2	0.25	0.17920	0.42455
λ_2	0.12	0.09735	0.01896

Remarks on EM approach

- EM approach applicable to general Phase-Type models

- Algorithm converges slowly (up to 10^4+ iterations).

- Occasionally converges to local maximum or saddle point.

- Method scales up linearly with sample size N:
must solve $N \cdot R \cdot (p+2)$ ODE's, where
$R = \#$ iterations and $p = \#$ transient states.

Numerical Results for Direct ML Estimates

- study performance of parameter estimates as function of sample size.

- special case : $(p, \mu, \lambda_1, \lambda_2) = (0.3, 2.0, 0.2, 0.3)$, and $(b_C, b_D, \beta_1, \beta_2)$ fixed $= (0, 0, 0, 0)$.

Parameter $(p, \mu, \lambda_1, \lambda_2)$ MLE's and SE's (in parentheses) on transformed scale (logit for p, log for others) by sample size N, for single simulated datasets with $k_1 = 4, k_2 = 3$.

	True	$N = 100$	$N = 1000$	$N = 10^4$	$N = 2 \cdot 10^4$	$N = 10^5$
logit(p)	-0.847	-0.427	0.169	-0.817	-0.935	-0.754
(SD)		(0.522)	(0.639)	(0.186)	(0.147)	(0.057)
log(μ)	0.693	-1.136	-0.983	0.457	0.730	0.578
(SD)		(0.906)	(1.026)	(0.232)	(0.188)	(0.082)
log(λ_1)	-1.609	-1.457	-1.398	-1.600	-1.623	-1.595
(SD)		(0.094)	(0.154)	(0.033)	(0.026)	(0.010)
log(λ_2)	-1.204	-0.696	-0.619	-1.184	-1.221	-1.177
(SD)		(0.279)	(0.347)	(0.046)	(0.031)	(0.015)

The Observed Fisher Information

- The per-observation Fisher Information matrices are estimated as $\frac{-\widehat{\mathbf{H}(\theta)}}{N}$; $\mathbf{H}(\theta)$ is the hessian matrix.

- This *observed Fisher Information* $\widehat{I}_1(\widehat{\vartheta})$ for $N = 10^5$ has eigenvalues 1.2601, 0.770, 0.0054, 0.0012

- Consider the linear combinations of the parameter estimates $v_1\,\text{logit}(\widehat{p}) + v_2\log(\widehat{\mu}) + v_3\log(\widehat{\lambda}_1) + v_4\log(\widehat{\lambda}_2))$, of eigenvectors of the Information matrix

- For large sample size N, theory predicts SE's $(1/\sqrt{\lambda_j N})$: $.891/\sqrt{N}$, $1.139/\sqrt{N}$, $13.550/\sqrt{N}$, $28.911/\sqrt{N}$

- 1st eigenvector linear combination of MLE's $.216\ \mathrm{logit}(\hat{p}) - .075\ \log(\hat{\mu}) - .425\ \log(\hat{\lambda}_1) - .877\ \log(\hat{\lambda}_2))$ is well estimated at 1.246 with predicted SE $= .028$.

- 4th eigenvector combination $.482\ \mathrm{logit}(\hat{p}) - .859\ \log(\hat{\mu}) + .082\ \log(\hat{\lambda}_1) + .152\ \log(\hat{\lambda}_2)$ is very badly estimated at 0.717 with predicted SE .914.

R Packages for Densities & Data Analysis

`actuar` *phase models* Goulet & Dutang computation of density involves numerical exp(tQ): ODE system

`dphtype, rphtype`, for phase–type density, simulation, & mgf

Parameter estimates in right-censored survival data via EM: Asmussen et al. (1996), Olsson (1996). **EMPht** C-program.

References

Aalen, Borgan & Gjessing, H. (2008) *Event History book*

Anderson, Chen, Jatoi & Rosenberg (2006) *Breast cancer data*

Armitage, P. & Doll, R. (1954, Br. J. Cancer)

Asmussen, Nerman, & Olssen (1996, Scand. J Stat) *EM*

Lee, M.-L. & Whitmore, G. (2006, Stat Sci) *threshold models*

Lee, R & Carter, L (1992 JASA) *mortality model*

Lin, X. & Liu, X. (2007, N Amer Actuar J) actuarial review

Moolgavkar, S. (2004 Int'l J Epidem) *re Multihit & cancer*

Neuts, M. (1981) *queueing phase models book*

Singpurwalla, N. (1996, IMS Lect.Notes) *risk & degradation*

Monte Carlo Results

Study asymptotic properties of the Hessian matrix by comparing estimated Fisher Information Matrices in two cases:

- Fisher Information matrix based on one iteration of 200000 simulated samples, $\hat{I}_1(\theta) = (-\widehat{H(\theta)})/200000$

- Fisher Information matrix based on B $(= 1000)$ iterations of 20000 simulated samples, $\hat{I}_2(\theta) = \dfrac{1}{B}\sum_{b=1}^{B}(-\widehat{H(\theta)}^{(b)})/20000$

- Results show well asymptotic property of Fisher Information matrix for large sample size.

Numerical Results

Fisher Information matrix based on one iteration of 200000 simulated samples (black) and B (= 1000) iterations of 20000 simulated samples (blue).

	logit(p)	log(μ)	log(β_1)	log(β_2)	log(λ_1)	log(λ_2)
logit(p)	0.0148	-0.0048	-0.0179	-0.0046	-0.0229	-0.0315
	0.0193	-0.0061	-0.0179	-0.0106	-0.0240	-0.0296
log(μ)	-0.0048	0.0950	0.0337	0.1011	0.0041	-0.0191
	-0.0061	0.1003	0.0316	0.1070	0.0038	-0.0213
log(β_1)	-0.0179	0.0337	0.0542	0.0859	0.0095	0.0065
	-0.0179	0.0316	0.0562	0.0733	0.0105	0.0101
log(β_2)	-0.0046	0.1011	0.0859	0.2160	-0.0430	-0.0621
	-0.0106	0.1070	0.0733	0.2203	-0.0411	-0.0631
log(λ_1)	-0.0229	0.0041	0.0095	-0.0430	0.1512	0.0099
	-0.0240	0.0038	0.0105	-0.0411	0.1644	0.0071
log(λ_2)	-0.0315	-0.0191	0.0065	-0.0621	0.0099	0.1349
	-0.0296	-0.0213	0.0101	-0.0631	0.0071	0.1337