Semi-parametric Survival Analysis with Time Dependent Covariates

Adam Branscum
University of Kentucky

Tim Hanson
University of Minnesota

and

Wesley Johnson
UC Irvine

Prakash Laud
Medical College of Wisconsin

Survival Analysis Background

- Cox Ph model is flexible but often fails to fit
- Semiparametric versions of AFT and proportional odds models are competitors
- Ancient attempts to provide semi-parametric approaches to the AFT model met with limited success (Miller, 1976; Buckley and James, 1979; Koul Susarla and Van Ryzin, 1981; Christensen and Johnson, 1988
- Recent Bayesian approaches are promising (Kuo and Mallick, 1997; Kottas and Gelfand, 2001; Gelfand and Kottas, 2002; Walker and Mallick, 1999, 2001; Hanson and Johnson, 2002, 2004.
- Recent frequentist approaches tend to focus on asymptotics for regression coefficients (Lin and Ying, 1995; Tseng, Wang and Hsieh, 2005)

Background

- Models/methods for Time Dependent Covariates are more sparse (Cox, 1972; Cox and Oakes, 1984; Robins and Tsiatis, 1992; Lin and Ying, 1995; Shyer et al., 1999)
- Frequentist joint modeling (Davidian and Tsiatis, 2003; Tseng, Hsieh and Wang, 2005)
- Bayesian approaches to joint modeling (Law, Taylor and Sandler, 2002; Brown and Ibrahim, 2003; Brown, Ibrahim and DeGruttola, 2005)
- We develop Bayesian semi-parametric approaches for the Cox, Cox and Oakes, and AFT models
- We also develop a Bayesian joint-modeling approach using Cox, Cox and Oakes and Proportional Odds models

The Basics of Survival Modeling

- Let $T > 0$ denote a random survival (event) time.
- $S(t) = P(T > t)$: Survival Function
- $h(t)dt = P(T \in [t, t + dt)|T \geq t)$: Hazard Function
- Denote risk factors as $x = (x_1, \ldots, x_p)$. The PH model relates covariates to the hazard and survival function as:

$$h(t|x) = \exp(x\beta)h_0(t)$$

$$S(t|x) = S_0(t)^{e^{x\beta}}$$

- Censored Survival Data:

$$\{t_i, \delta_i, x_i : i = 1, ..., n\}$$

Alternative Models

- **AFT Model:**

$$S(t|x) = S_0(\exp(x\beta)t) \Leftrightarrow T = \exp(x\beta)V \ :$$

- **Prop. Odds Model:**

$$\frac{S(t|x)}{1 - S(t|x)} = e^{x\beta}\frac{S_0(t)}{1 - S_0(t)}$$

Models for S_0

- Dirichlet Process (DP) (Ferguson, 1973)
- Mixtures of Dirichlet Processes (MDP) (Antoniak, 1974)
- Dirichlet Process Mixtures (DPM) (Escobar, 1994)
- Polya Tree and Mixture of PT Process (MPT) (Lavine, 1992, 1994; Hanson, 2006)

- Alternative to semi-parametric models

  Dependent Dirichlet Process (DDP) regression model (MacEachern, 1999: De Iorio et al. 2004; De Iorio et al. 2007)

Polya Trees

- Split sample space Ω into two disjoint sets B_0 and B_1; further split B_0 into B_{00} and B_{01}, split B_1 into B_{10} and B_{11}:

B_0		B_1	
B_{00}	B_{01}	B_{10}	B_{11}

- Define

$$Y_0 = P(V \in B_0), \qquad Y_1 = P(V \in B_1),$$

$$Y_{00} = P(V \in B_{00} | V \in B_0),$$

$$Y_{01} = P(V \in B_{01} | V \in B_0),$$

$$Y_{10} = P(V \in B_{10} | V \in B_1),$$

$$Y_{11} = P(V \in B_{11} | V \in B_1).$$

- Then $P(V \in B_{ij}) = Y_i Y_{ij}$.

- Continue: Let $\epsilon = \epsilon_1 \cdots \epsilon_m$ be an arbitrary binary number.
- Split $B_\epsilon \longrightarrow \{B_{\epsilon 0}, B_{\epsilon 1}\} \qquad \forall \epsilon$.
- Then

$$\left.\begin{array}{l} Y_{\epsilon 0} = P(V \in B_{\epsilon 0} | V \in B_\epsilon) \\ Y_{\epsilon 1} = P(V \in B_{\epsilon 1} | V \in B_\epsilon) \end{array}\right\} \Rightarrow$$

$$P(V \in B_{\epsilon_1 \cdots \epsilon_m}) = \prod_{j=1}^{m} Y_{\epsilon_1 \cdots \epsilon_j}$$

PT

- Create random PM on S_0:

$$(Y_{\epsilon_1\cdots\epsilon_m 0}, Y_{\epsilon_1\cdots\epsilon_m 1}) \sim \text{Dir}(\alpha_{\epsilon_1\cdots\epsilon_m 0}, \alpha_{\epsilon_1\cdots\epsilon_m 1})$$

- Random S_0 specified by
- $\Pi = \cup_{j=1}^{\infty}\{B_{\epsilon_1\cdots\epsilon_j} : \epsilon_1 \cdots \epsilon_j \in \{0,1\}^j\}$
- $\mathcal{A} = \cup_{j=1}^{\infty}\{\alpha_{\epsilon_1\cdots\epsilon_j} : \epsilon_1 \cdots \epsilon_j \in \{0,1\}^j\}$

PT

- $S_0|\Pi, \mathcal{A} \sim PT(\Pi, \mathcal{A})$
- Lavine (1992, 1994) catalogues Polya tree theory
- Conjugacy: $V_1|S_0 \sim S_0 \longrightarrow$

$$S_0|V_1, \Pi, \mathcal{A} \sim PT(\Pi, \mathcal{A}^*),\ \mathcal{A}^* = \{\alpha_\epsilon + I_{B_\epsilon}(V_1)\}$$

- Specify Π and $\mathcal{A}$ only to level $M \longrightarrow$
  "partially specified Polya tree"
- $S_0|\Pi_M, \mathcal{A}_M \sim FPT(\Pi_M, \mathcal{A}_M)$

PT

- Ferguson (1974):

  $$\alpha_{\epsilon_1 \cdots \epsilon_{m-1} 0} = \alpha_{\epsilon_1 \cdots \epsilon_{m-1} 1} = cm^2$$

  $$\Rightarrow \qquad S_0 \quad \text{absolutely continuous}$$

- Large c results in a parametric analysis, and small c results in a more non-parametric analysis

Center Process Around S_θ

- By definition of the process

$$E\{S_\theta(B_{\epsilon_1\cdots\epsilon_m})\} = \left(\frac{\alpha_{\epsilon_1}}{\alpha_0+\alpha_1}\right)\left(\frac{\alpha_{\epsilon_1\epsilon_2}}{\alpha_{\epsilon_1 0}+\alpha_{\epsilon_1 1}}\right)\cdots \cdots\left(\frac{\alpha_{\epsilon_1\cdots\epsilon_m}}{\alpha_{\epsilon_1\cdots\epsilon_{m-1}0}+\alpha_{\epsilon_1\cdots\epsilon_{m-1}1}}\right)$$

- If $\alpha_{\epsilon 0} = \alpha_{\epsilon 1}$ for all ϵ, then $E\{S_\theta(B_{\epsilon_1\cdots\epsilon_m})\} = 2^{-m}$.
- $S_\theta(B_{\epsilon_1\cdots\epsilon_m}) = 2^{-m} \Rightarrow E\{S_0(B_\epsilon)\} = S_\theta(B_\epsilon)$

Predictive Density

- Let

$$V_i \sim S_0,\ i = 1, ...n + 1$$

$$S_0|\Pi, \mathcal{A} \sim PT(\Pi, \mathcal{A})$$

$$V = (V_1, \ldots, V_n)'$$

- Define $f_\theta = -S'_\theta$

Pred Dens and Marg Post for β

$$f_{V_{n+1}}(w|V) =$$

$$\left\{ \prod_{j=2}^{M(w)} \frac{cj^2 + n_{\epsilon_j(w)}(V)}{2cj^2 + n_{\epsilon_{j-1}(w)}(V)} \right\} 2^{M(w)-1} f_\theta(w),$$

For the AFT model

$$p(\beta|data) \propto p(\beta) \times$$

$$\prod_{j=1}^{n} f_{V_j}(T_j e^{-x_j\beta} | V_i = T_i e^{-x_i\beta}, i < j) e^{-x_j\beta}$$

Mixture of Polya Trees

- Can make exact inferences for β
- However, choosing S_θ for particular fixed θ, is *ad hoc* & the partition affects inferences for β
- Solution: Mixture of Polya Trees

$$S_0 | \Pi^\theta, \mathcal{A} \sim PT(\Pi^\theta, \mathcal{A})$$

$$\theta \sim p(\theta), \quad \beta \sim p(\beta)$$

- Has the additional nice property of centering on a parametric family, like the family of log normal pdf's, or Weibull family...

Full AFT Model with MPT Prior

$$T_i = exp(x_i \beta) V_i$$

$$V_1, \ldots, V_n | S_0 \stackrel{iid}{\sim} S_0, \quad S_0 | \theta \sim PT(\Pi^\theta, \mathcal{A})$$

$$\beta \sim p(\beta), \quad \theta \sim p(\theta)$$

- Predictive density for $T_{n+1}|x, data$ is differentiable everywhere; partition effects are "smoothed"
- Exact inference for $\beta, \theta | data$ is possible
- S_0 centered on a parametric family of probability distributions
- We set $S_0(0, 1] = 0.5$ with probability one
- Results in median regression eg. $med(T)|x = e^{x\beta}$

- Can place prior on c
- Easy to incorporate informative prior information for β as in BCJ (1999) or Ibrahim and Chen (2000)
- Can use output from parametric analysis in constructing candidate in Metropolis sampler

Time Dependent Covariates

- **Stanford Heart Transplant Data:** Time of HTP is not known at the beginning of the study.
- Let $Z_1(t)$ be zero until the time of HTP and one afterwards
- Let $Z_2(t)$ be the mismatch score between donor and recipient hearts. Takes the value zero before HTP and a particular value afterwards
- Goal is to measure effect of HTP and mismatch score on survival prospects.

Cerebral Edema (CE)

- CE is a complication of diabetic ketoacidosis (DK) in children
- Children are admitted to the hospital for DK and CE may or may not occur
- Children are monitored over time. The response is time to CE after entry into the hospital
- Fixed covariates are age and BUN
- Time Dependent covariates are Sodium administered, fluids administered, and bicarbonate administered
- Goal is to determine if procedures of administering various fluids is hastening the onset of CE

Cox TDC Model (CTD)

- Let $\{z(t) : t > 0\}$ be a vector of TDC covariate processes, which we assume are fixed and known for now
- Define the Cox TD hazard function as

$$h(t|z, \beta) = e^{z(t)\beta} h_0(t)$$

  where $h_0(\cdot)$ is an arbitrary "baseline" hazard function
- Let $\{r_j, j = 0, 1, \ldots\}$ be the grid of times over which $z(t) : t > 0$ is constant, eg. no known changes
- Denote the r_j's as changepoints for the covariate process
- Relative hazard for any two individuals is constant in between each adjacent pair of changepoints

AFT TDC Model (AFTD)

- Prentice and Kalbfleisch (1979)

$$h(t|z,\beta) = e^{z(t)\beta} h_0(te^{z(t)\beta})$$

- Can show that this model is equivalent to a mixture of truncated AFT models over each of the adjacent changepoint intervals, $([r_{j-1}, r_j))$, where the acceleration factor (AF) for the jth interval is $c_j \equiv e^{z(r_{j-1})\beta}$.
- Both the CTD and AFTD models presume that the risk of failure at time t only depends on the current values of the TDC's, and not their history.

Cox and Oakes TDC Model (COTD)

- Model assumes that an individual with covariate $z(\cdot)$ uses up their time at a rate of $e^{z(t)\beta}$ relative to "baseline", namely

$$T_0 = \int_0^T e^{z(s)\beta} ds.$$

- The corresponding hazard function is

$$h(t|z,\beta) = e^{z(t)\beta} h_0(\bar{c}(t)t),\ \bar{c}(t) = \frac{1}{t}\int_0^t e^{z(s)\beta} ds$$

- This model presumes that there is a cumulative effect of the covariate process up to time t that will effect the hazard of failure at that time.

MFPT Baseline for All Models

- Assume the same MFPT prior for *all* three models, eg.

$$S_0 \sim PT(A_M, \Pi_M^\theta),\ \theta \sim p(\theta)$$

- Center PT on the family $\{S_\theta : \theta \in \Theta\}$
- Assume that, for given θ, the prior on the intervals at the highest level of the tree is governed by S_θ
- A Lik cont (no marg) for the AFTD model is $L_z(\beta, \Xi_M, \theta | T = t) =$

$$\left\{ \prod_{j=1}^{m} p_j \right\} \frac{2^M p_N^\theta f_\theta(c_{m+1} t) c_{m+1}}{S_\theta(c_{m+1} r_m | \Xi_M)},$$

where $p_j = S_0(c_j r_j | \Xi_M)/S_0(c_j r_{j-1} | \Xi_M)$

Likelihood Functions

- The likelihood contribution for an observation right-censored at time t is $L_z(\beta, \Xi_M, \theta | T > t) =$

$$\left\{ \prod_{j=1}^{m} p_j \right\} \frac{S_0(c_{m+1} t | \Xi_M, \theta)}{S_0(c_{m+1} r_m | \Xi_M, \theta)}$$

- The complete data involve n independent event times, $\{t_i\}_{i=1}^n$, that are the observed survival times $(T_i = t_i)$ or are right-censoring times $(T_i > t_i)$, and
- n covariate processes $\{z_i(\cdot)\}_{i=1}^n$
- The complete likelihood is

$$L(\beta, \Xi_M, \theta) = \prod_{i=1}^{n} L_i(\beta, \Xi_M, \theta)$$

Gibbs Sampling

- Alternate between sampling $\beta, \theta | \Xi_M$ and $\Xi_M | \beta, \theta$
- The former can be sampled via Metropolis-Hastings using a parametric model in WinBUGS or SAS to obtain a suitable candidate distribution
- Use MH for updating the components $(Y_{\epsilon 0}, Y_{\epsilon 1})$, with candidate

  $$(Y^*_{\epsilon 0}, Y^*_{\epsilon 1}) \sim \text{Beta}(mY_{\epsilon 0}, mY_{\epsilon 1})$$

  typically $m = 20$ or 30
- Can easily handle interval censored data
- Other likelihoods are similarly obtained

Simulated Data

- Simulate data from true baseline of log normal(0.69, 0.04) with two distinct TDC's
- The first TDC is constant at zero, and the second is zero up to one unit of time and is one thereafter.
- Ten data points with TDC 1 and 90 with TDC 2
- The regression coefficient is $\beta = 0.69$
- Fit MFPT with $c = 1$ and $M = 4$, and with log-logistic family as base
- Uniform priors on finite intervals for $(\theta_1, \theta_2, \beta)$

	AFTD	CO	PH
$E(\ell n(\text{Lik}))$	55	47	49.5
LPML	51	42	46
β	.65	1.73	3.17
Prob Interval	(.48,.96)	(1.34,2.22)	(2.23,4.22)

Posterior inferences for simulated data.

Candidate Generating Distributions

- If S_θ is exponential with parameter θ, then the AFTD, COTD, and CTD models are the same
- The likelihood is

$$L(\beta, \theta) =$$

$$\prod_{i=1}^{n} \left\{ \prod_{j=1}^{J_i} e^{-\theta[r_{ij} - r_{i,j-1}] e^{x_{i,j-1}\beta}} \right\} e^{[t_i - r_{i,J_i}] e^{-\theta x_{i,J_i}\beta}} \theta^{\delta_i}$$

- Readily implemented in SAS, S-plus, WinBUGS... to obtain starting values and covariance matrices for the candidate generating distribution (CGD)

CGD's

- We generally used the log-logistic to center the three MPT survival models
- Used WinBUGS fit to get rough candidate generating covariance matrix for (β, θ) using random-walk M-H chain
- Only needed 10,000 iterates in the final runs. Can all be easily automated
- Jara, DP Package

Stanford Heart Transplant Data

- Data on patients admitted to Stanford Program and analyzed using the Cox model with TDC's (Crowley and Hu, 1977)
- Lin and Ying (1995) use same data to illustrate their heuristic procedure for COTD justified by asymptotic properties
- We fit data using CTD, COTD and AFTD models with MFPT prior; $M = 5$ and $c = 1$.

Stanford Study

$$x_{i1}(t) = \begin{cases} 0 & \text{if} \quad t < z_i \\ 1 & \text{if} \quad t \geq z_i \end{cases}$$

$$x_{i2}(t) = \begin{cases} 0 & \text{if} \quad t < z_i \\ \text{age at transplant} - 35 & \text{if} \quad t \geq z_i \end{cases}$$

$$x_{i3}(t) = \begin{cases} 0 & \text{if} \quad t < z_i \\ \text{mismatch score} - 0.5 & \text{if} \quad t \geq z_i \end{cases}$$

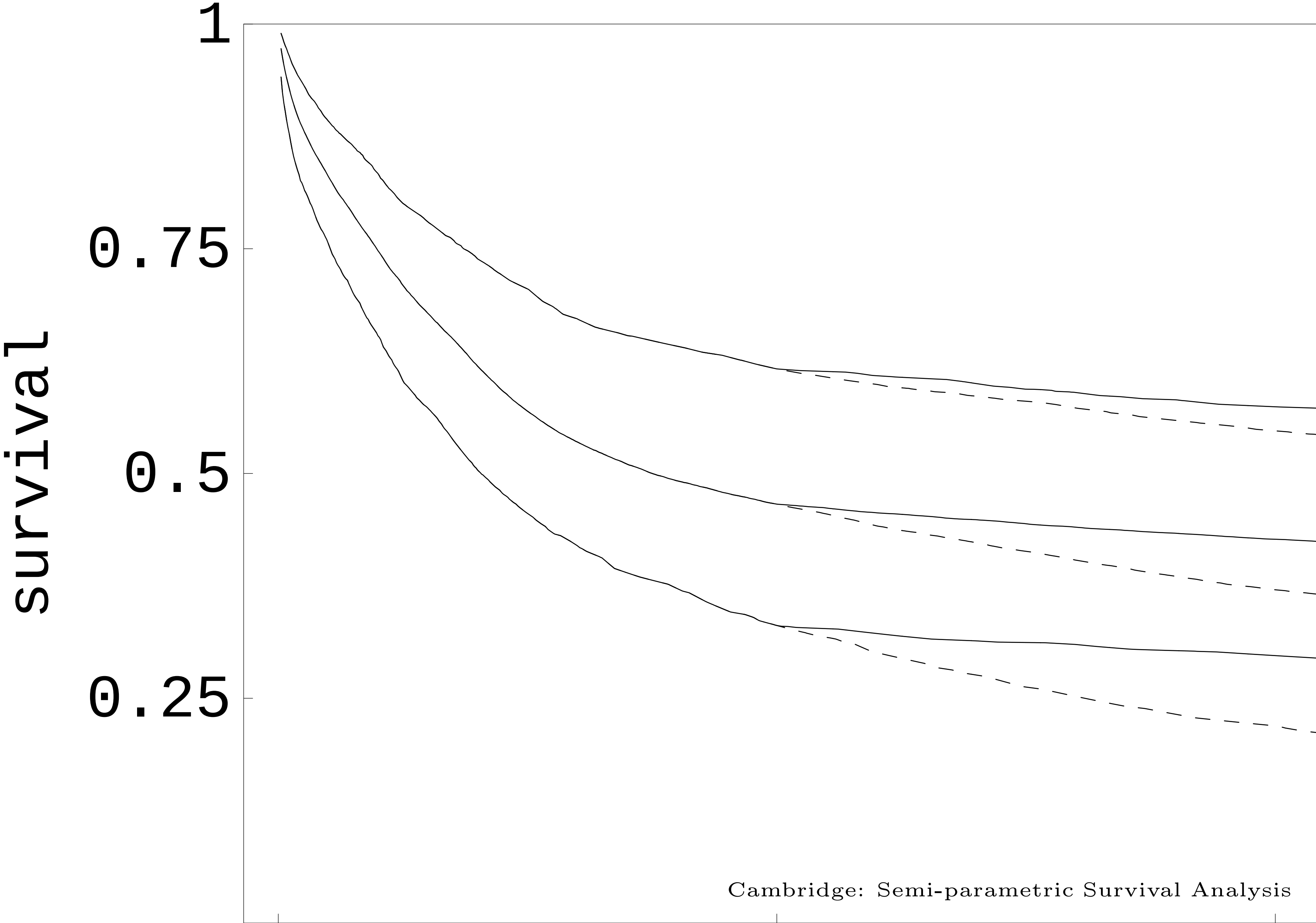
Stanford Study

	AFTD	COTD	CTD
ELL	-461	-460	-458
$LPML$	-468	-467	-464
Stat	-1.76	-1.10	-1.04
	(-3.86,1.57)	(-2.70,0.50)	(-1.99,-0.17)
Age-35	0.104	0.054	0.058
	(-0.020,0.260)	(-0.004,0.133)	(0.015,0.107)
Mis-0.5	1.63	0.64	0.49
	(-0.38,3.89)	(-0.30,1.52)	(-0.09,1.03)

Stanford Study

- The relative hazard (RH), comparing individual w/ no HTP to an individual how gets one after 6 months

$$2.83 \quad (1.19, 7.31)$$



1
0.75
0.5
0.25
survival
Cambridge: Semi-parametric Survival Analysis

Stanford Study

- Parametric exponential yielded posterior median estimates for $(\beta_1, \beta_2, \beta_3)$

$$(-2.74, 0.08, 0.98)$$

$$LPML = -486.3$$

- Integrated Cox-Snell residuals show extreme curvature
- Lin and Ying (1995) semiparametric-partial-likelihood estimates

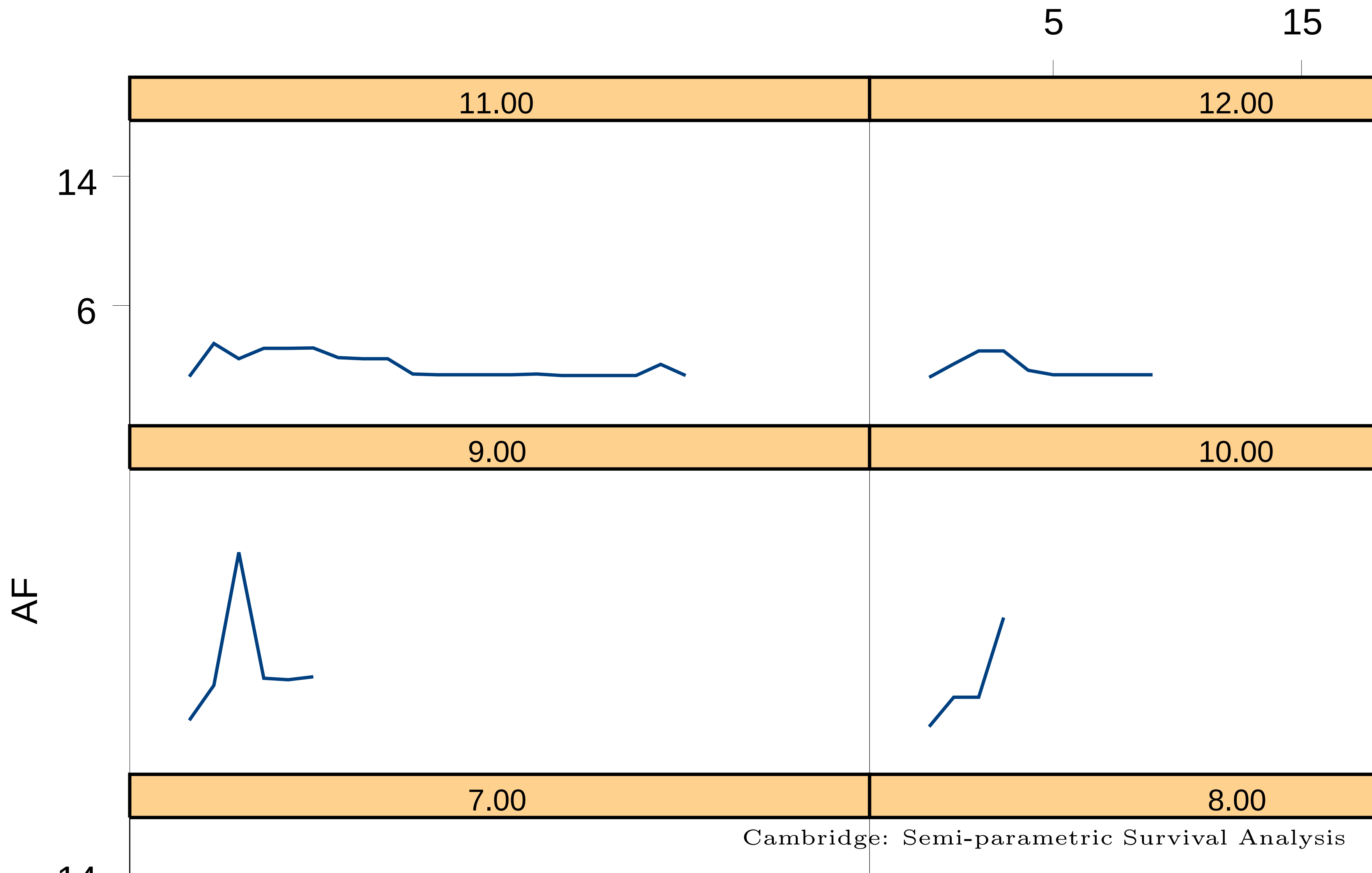
$$(-1.99, 0.096, 0.93)$$

Closer to exponential than semiparametric

CE Data

- Range of LPML's ranged between -175 to - 176
- AFTD appears to fit the best based on residual plots

Relative hazards in the OR over time



Joint modeling setting

- Longitudinal data associated with terminal event of interest
- Conditional on longitudinal process, we have survival analysis with TDC's
- Longitudinal process is often observed with error
- With TDC's, process was assumed constant between observation times
- Can lead to bias (Prentice, 1982)
- Joint modeling is used to make inferences for assessing:
  1. Trends in the time course of a longitudinal process
  2. The association between de-noised time-dependent processes and event prognosis

Alternatives to Joint Modeling

- Don't model the longitudinal data. Survival analysis with TDCs (subsequently called RAW)
- Two-stage procedures (called Imputation):

  Model the observed longitudinal process assuming it has noise

  Impute the de-noised signal process; treat it as a TDC
- Compare joint analyses with these

Joint Modeling

- Model the longitudinal data

$$f(y(\cdot)|\gamma)$$

Conditional on that, model the survival time,

$$f(T|y, \xi)$$

- Longitudinal process, $x_i(\cdot)$, is measured with error so we observe $y_i(\cdot)$ at several time points where

$$\begin{aligned} y_i(t) &= x_i(t) + \epsilon_i(t) \\ x_i(t) &= \mathbf{f}(t)\boldsymbol{\gamma} + \mathbf{g}(t)\mathbf{b}_i + U_i(t) + \mathbf{z}_i\boldsymbol{\alpha} \\ \epsilon_i(t) &\overset{iid}{\sim} N(0, \sigma^2) \end{aligned}$$

Imputation

- Use the longitudinal model to obtain $\hat{x}_i(t)$
- Use data $\{t_i, \delta_i, \hat{x}_i\}$ as if $\hat{x}_i$ were observed

- Define the cumulative history $X_t = \{x(s) : s \leq t\}$

Inferences: Bayesian Joint Modeling

- Here, (after some modeling) we obtain,

$$f(y_f, T_f|\text{data}) = f(y_f|\text{data}) f(T_f|y_f, \text{data})$$

  y_f is a hypothetical observed history

- Prognosis based on their predictive density, $f(T_f|y_f, \text{data})$. Compare these for different hypothetical histories. Set $y_f = y_i$

- Conditional hazards:

$$h(t|X_t, \text{data}) = \int h(t|X_t, \xi) p(d\xi|\text{data})$$

  for hypothetical X_t.

Models for survival data with TDC's

- Tseng et al (2005) developed a semiparametric frequentist joint model using the COAFT (Monte Carlo EM algorithm with bootstrap se's for reg coeffs)

- Sundaram (2006) extended the proportional odds model to allow for TDCs yielding a POTDC model, which is defined by

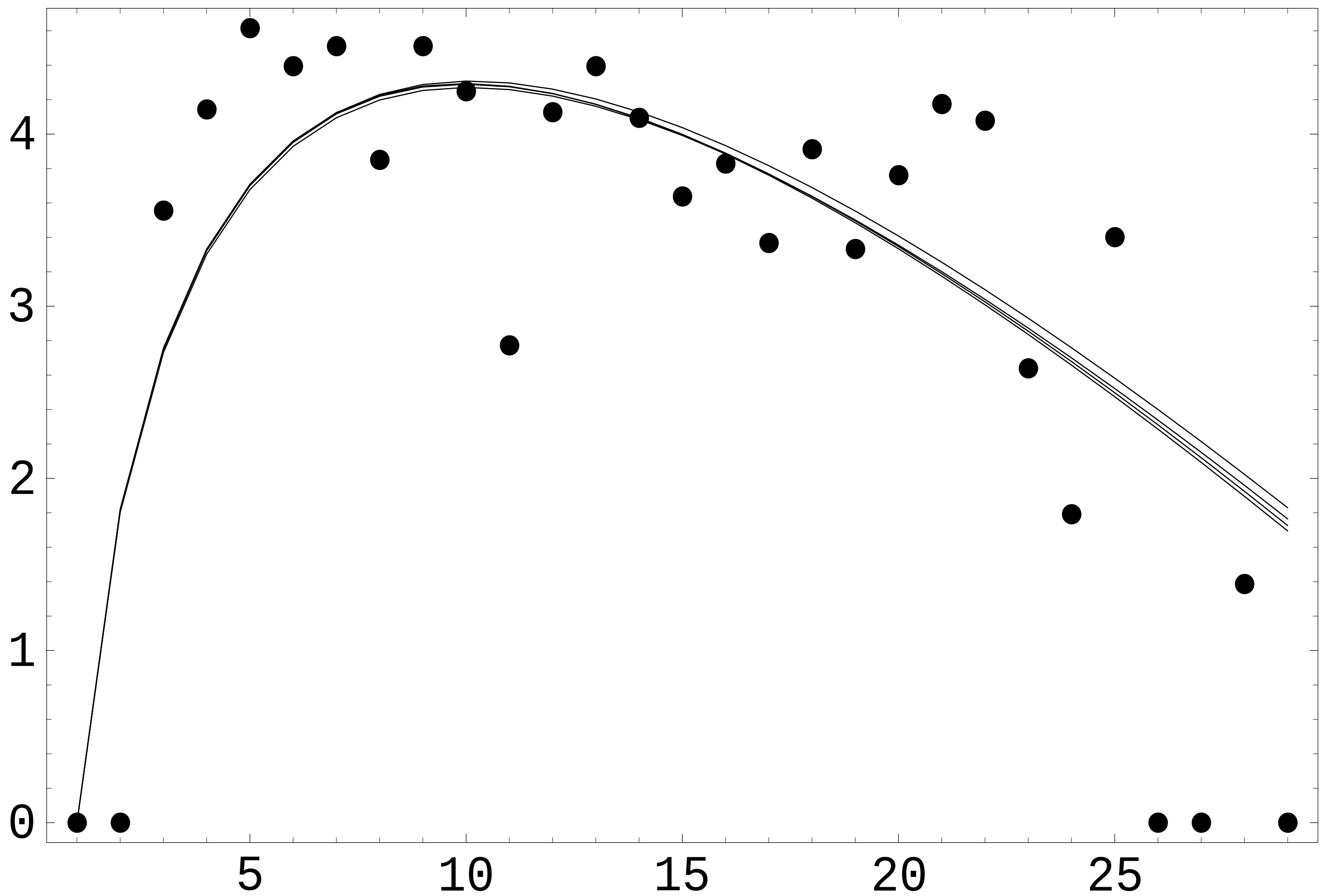
$$\frac{d}{dt}\left[\frac{1 - S(t|x(\cdot))}{S(t|x(\cdot))}\right] = e^{x(t)\beta}\frac{d}{dt}\left[\frac{1 - S_0(t)}{S_0(t)}\right]$$

Illustration: Medfly Data

- Data from a study on reproductive patterns of 1000 female Mediterranean fruit flies.
- Obtained by recording the number of eggs produced each day throughout their lifespans
- Goal was to examine the association between egg production patterns and lifetime
- Sample size of 251 flies with lifespans ranging from 22 to 99 days, and no censored observations

Fitted trajectory: Fly 1

- Fitted trajectory for a "typical" medfly. Similar shapes for PO, PH, CO, and longitudinal only analysis



Model for longitudinal data

- Compare with a previous joint analysis (Tseng et al, 2005), *so we use their structure*
- $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})'$ are the n_i longitudinal measurements of subject i at times $\mathbf{t}_i = (t_{i1}, \ldots, t_{in_i})'$
- Model specifies that trajectories satisfy

$$y_{ij} | \mathbf{b}_i, \sigma^2 \overset{\perp}{\sim} N\left(b_{i1} g_1(t_{ij}) + b_{i2} g_2(t_{ij}) + \cdots + b_{id} g_d(t_{ij}), \sigma^2\right)$$

- Individual trajectories

$$\mathbf{b}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma} \overset{iid}{\sim} N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Model fitting

- Let $x_i(t|\mathbf{b}_i) = b_{i1}g_1(t) + \cdots + b_{id}g_d(t)$
- For joint models, survival is specified conditional on

$$\{x_i(\cdot|\mathbf{b}_i)\}_{i=1}^n$$

- S_0 modeled with MFPT prior
  - log-logistic centering family, i.e. $E(S_0(t)) = (1 + t^{1/\tau}e^{-\alpha/\tau})^{-1}$
  - collection of branch probabilities Ξ_M
  - weight parameter c
- Let $\boldsymbol{\theta} = (\alpha, \tau, \Xi_M, c)$
- A model $[T_i|\boldsymbol{\theta}, \beta, x_i(\cdot|\mathbf{b}_i)]$ is specified as CO, PO, or PH

Model fitting

- Independent priors:
  - $p(\mu, \mathbf{\Sigma}, \beta, \alpha, \tau) \propto |\mathbf{\Sigma}|^{-(d+1)/2}$
  - $p(\sigma^{-2}) \propto 1/\sigma^{-2}$
  - $c \sim \Gamma(c|a_c, b_c)$
  - $(X_{j,2k-1}, X_{j,2k}) \sim \text{Dirichlet}(cj^2, cj^2)$
- The posterior based on the survival portion, the longitudinal portion, and the prior is then

$$p(\beta, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{\Sigma}, \sigma|\mathbf{T}, \mathbf{y}_{1:n}) =$$

$$\left[\prod_{i=1}^{n} f(T_i|x_i(\cdot|\mathbf{b}_i), \boldsymbol{\theta}, \beta)^{\delta_i} S(T_i|x_i(\cdot|\mathbf{b}_i), \boldsymbol{\theta}, \beta)^{1-\delta_i}\right] \times$$

$$\left[\prod_{i=1}^{n} p(\mathbf{y}_i|\mathbf{b}_i, \sigma) p(\mathbf{b}_i|\boldsymbol{\mu}, \mathbf{\Sigma})\right] p(\beta, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{\Sigma}, \sigma)$$

Model fitting

- The full conditional distributions for $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and σ^{-2} are:

$$
\begin{aligned}
\boldsymbol{\Sigma}^{-1}|\mathbf{b}_{1:n}, \boldsymbol{\mu} &\sim \text{Wishart}\left(n, \left[\sum_{i=1}^{n}(\mathbf{b}_i - \boldsymbol{\mu})(\mathbf{b}_i - \boldsymbol{\mu})'\right]^{-1}\right) \\
\boldsymbol{\mu}|\mathbf{b}_{1:n}, \boldsymbol{\Sigma} &\sim N_d\left(\bar{\mathbf{b}}_\bullet, \boldsymbol{\Sigma}/n\right) \\
\sigma^{-2}|\mathbf{b}_{1:n} &\sim \Gamma\left(0.5\sum_{i=1}^{n} n_i, 0.5\sum_{i,j}(y_{ij} - x_i(t_{ij}|\mathbf{b}_i))^2\right)
\end{aligned}
$$

- Metropolis-Hastings steps were used to sample the full conditionals for the $\mathbf{b}_i$'s (random-walk M-H), Ξ_M (w/ beta proposals), c (w/ truncated normal proposal), (α, β, τ) (w/ random walk M-H).

Illustration: Medfly Data

Response

$$\ln(y_i(t) + 1)$$

and

$$x_i(t|\mathbf{b}_i) = b_{1i}\ln(t) + b_{2i}(t-1)$$

Model comparison

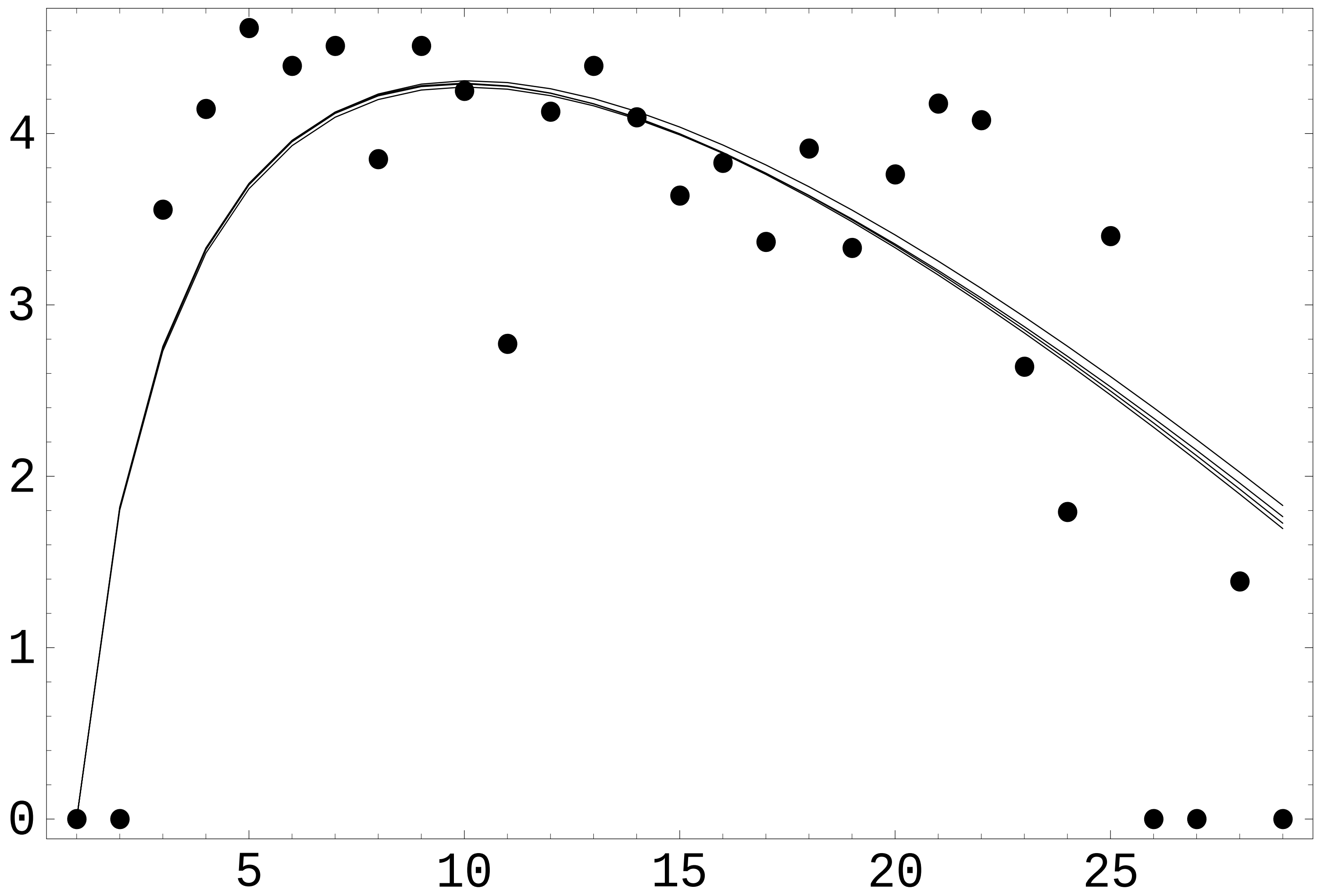
- *negative*-LPML statistics (smaller is better) comparing modeling approaches:

Model	Method	PO	PH	CO
parametric	raw	867	870	937
MPT	raw	865	866	938
MPT	imputed	947	959	973
parametric	joint	947	959	973
MFPT	joint	945	956	973

- Summary based on LPML criterion:
  - Predictively, PO and PH models preferred over CO
  - Survival with fixed TDC's preferred over joint
  - MFPT improves predictive performance *only slightly* compared to parametric model

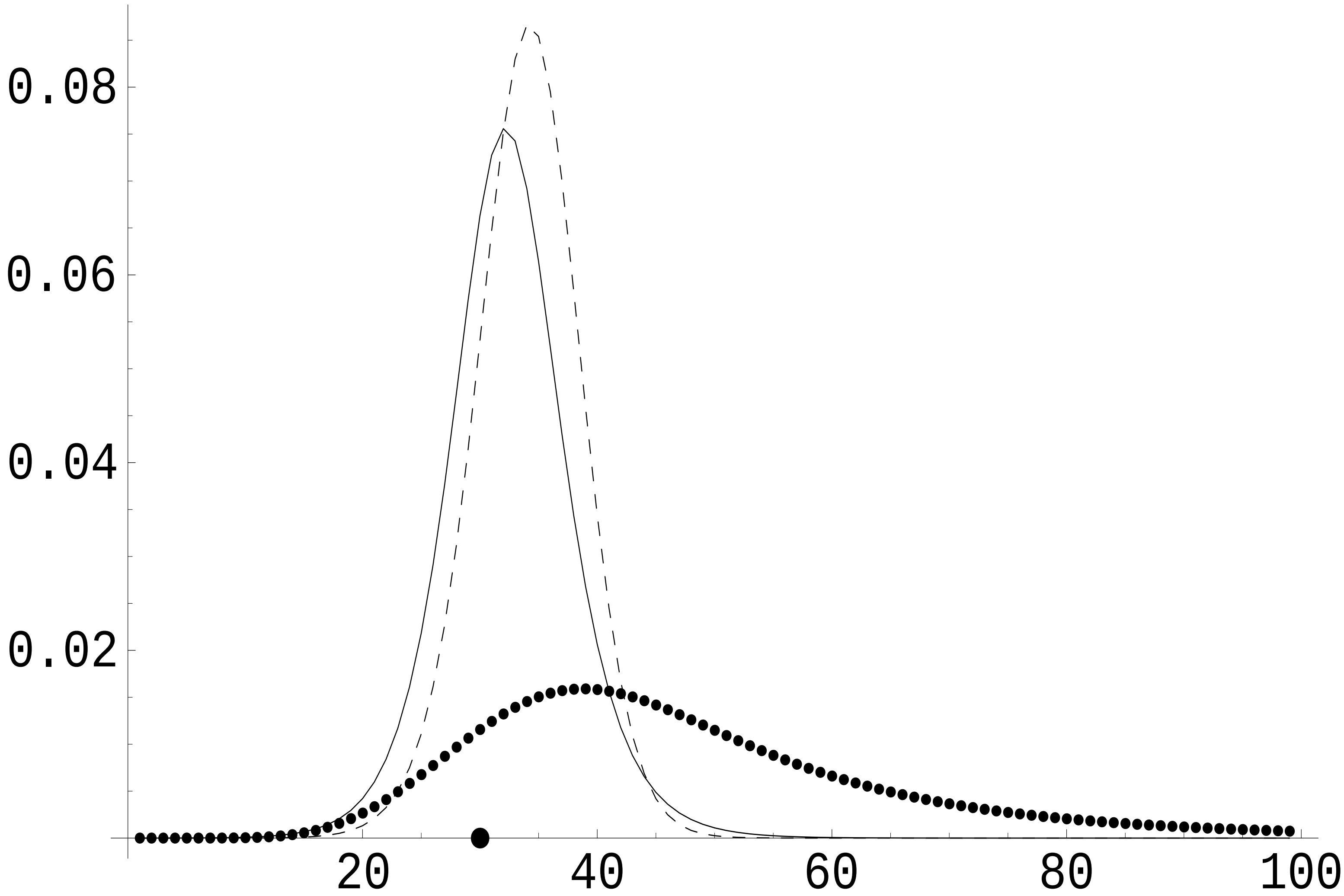
Fitted trajectory: Fly 1

- Fitted trajectory for a "typical" medfly. Similar shapes for PO, PH, CO, and longitudinal only analysis



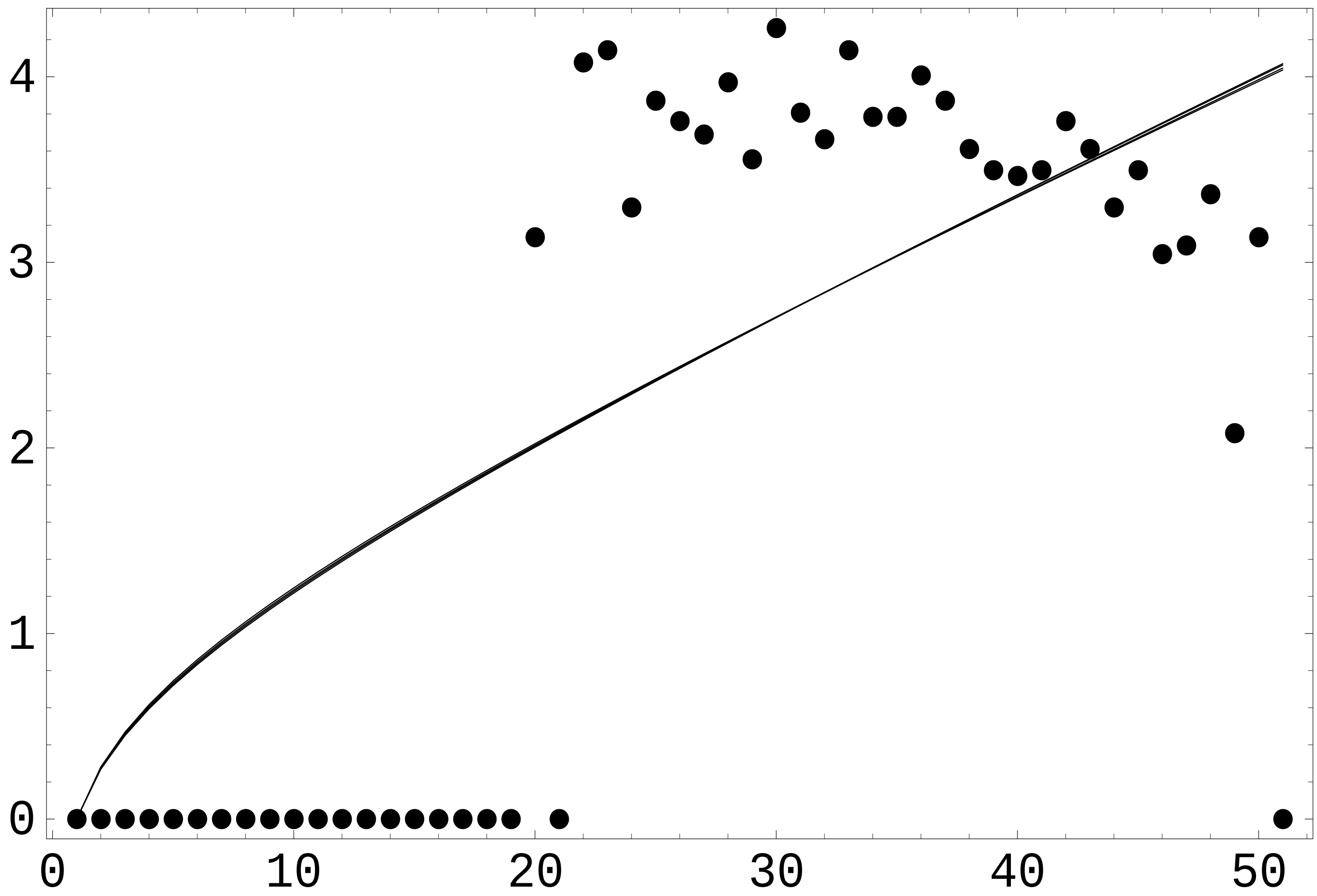
Predictive survival density: Fly 1

- Solid is PO, dashed is PH, and dotted is CO



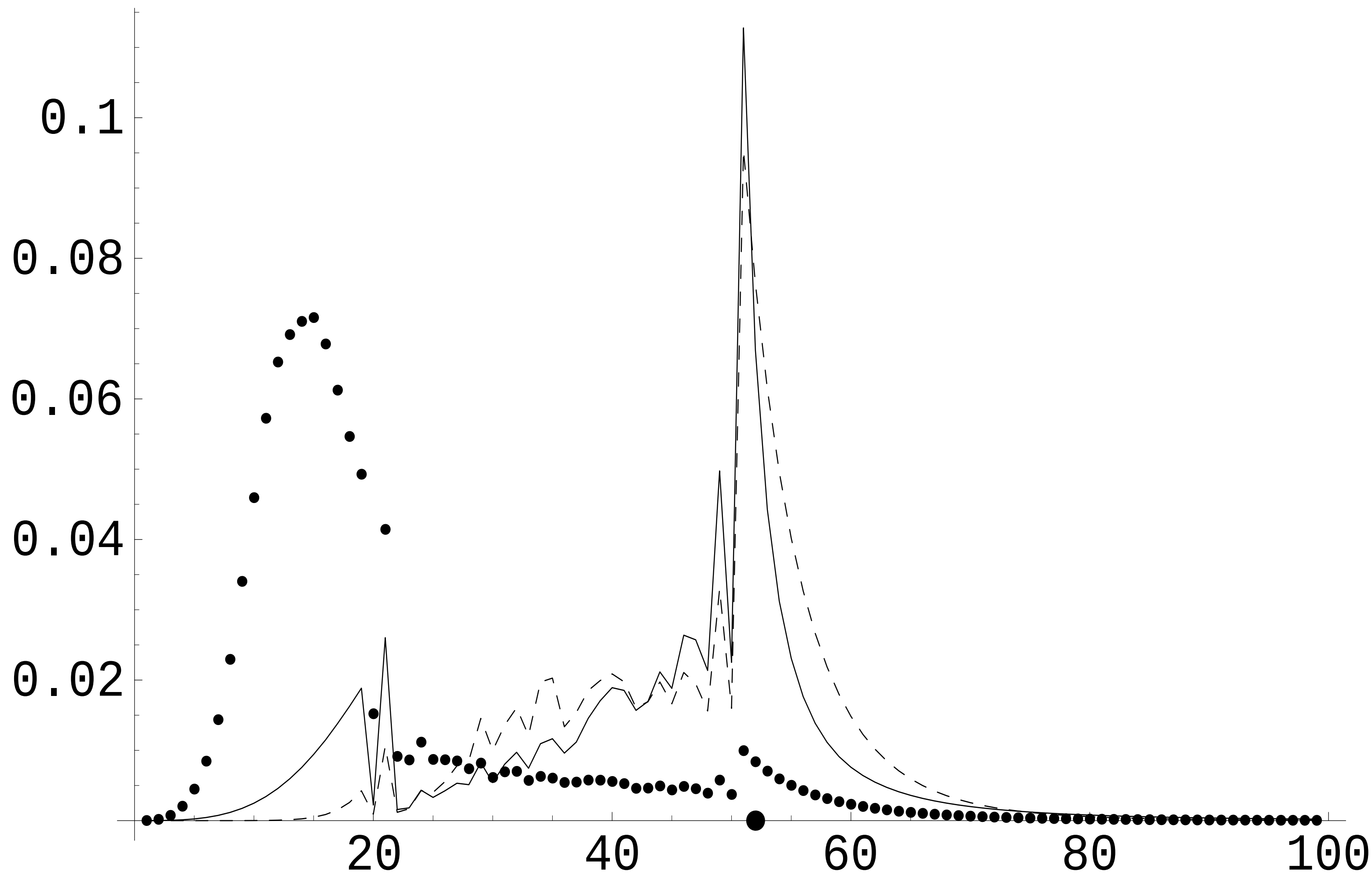
Fitted trajectory: Fly 2

- Fitted trajectory for another medfly using PO, PH, CO, and longitudinal only analysis



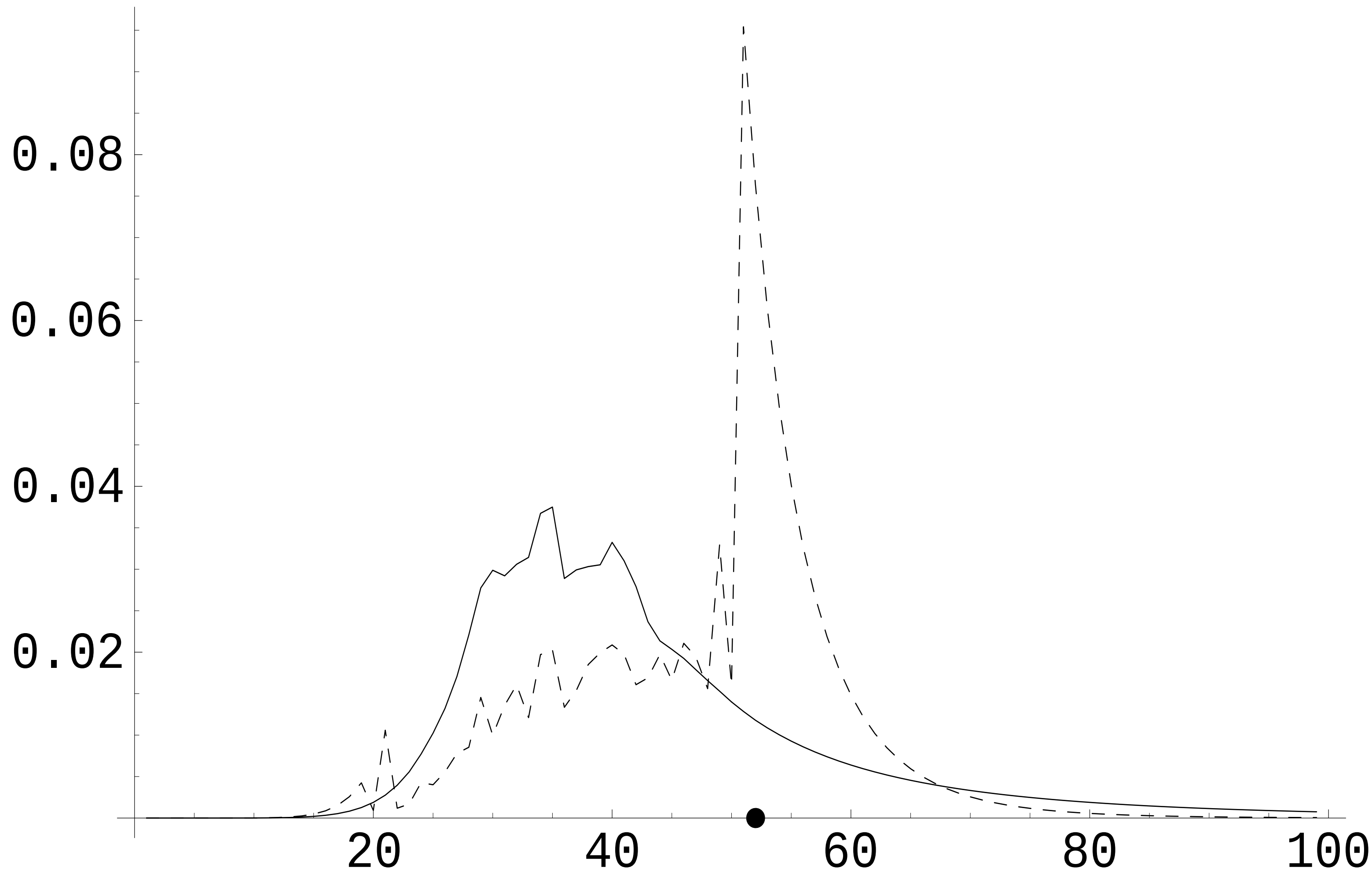
Predictive survival density: Fly 2

- PO (solid), PH (dashed) and CO (dotted) analyses using **Raw** trajectories.



Predictive survival density: Fly 2

- Raw trajectory (dashed line); **joint** analysis (solid line)



Posterior inference for β

Model	Method	PO	PH	CO
parametric	raw	-0.75	-0.65	-0.36 (-0.44,-0.27)
MPT	raw	-0.74	-0.64	-0.37 (-0.45,-0.29)
MPT	imputed	-0.74	-0.37	0.16 (-0.01,0.30)
parametric	joint	-0.78	-0.39	0.19 (0.01,0.33)
MPT	joint	-0.79	-0.40	0.19 (0.01,0.32)

- $\Pr(\beta < 0 | \mathbf{T}, \mathbf{y}_{1:n}) = 1$ for PO and PH models
  $\Rightarrow$ survival prospects are better for the most fertile flies.
- Inferences based on CO are different for joint models than for models based on raw trajectories

Why I like MFPT's for SA

- Prior centered on parametric family; DPM Not
- Easy to place informative prior on reg coeffs; DPM Not
- No need to marginalize over S_0
- Inferences on functionals of S_0 simple
- Median regression is immediate; DPM not
- No "sticky clusters"
- Hanson (2006, JASA)
- Hanson, T., Branscum, A., and Johnson, W.O. (2005). Bayesian nonparametric modeling and data analysis: an introduction. In Bayesian Thinking: Modeling and Computation (Handbook of Statistics, volume 25)