The AFT model framework
Estimation and inference
survreg

Accelerated Failure Time Models

Patrick Breheny

October 11

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

Introduction

- Last time, we introduced the Weibull distribution and saw that, on the log scale, it could be viewed, essentially, as a regular linear regression model, albeit with extreme value residual terms

- Today we will look at this modeling framework in more detail, discuss estimation and model fitting, and go through an example in which we fit the model to real data

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

Accelerated failure time models

- For a random time-to-event T, an *accelerated failure time* (AFT) model proposes the following relationship between covariates and $Y = \log T$:

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + W_i,$$

where $W_i \overset{\perp\!\!\!\perp}{\sim} f$ are the error, or residual, terms; such models are also sometimes referred to as *log-linear models*
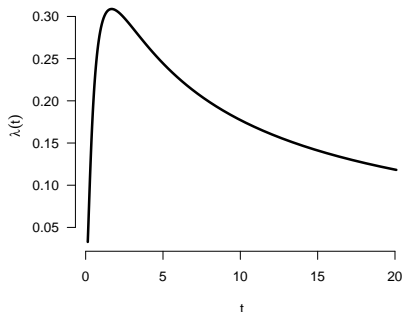
- The above framework describes a general class of models: depending on the distribution we specify for W, we will obtain a different model, but all will have the same general structure

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

The Lognormal AFT

- For example, an obvious possibility is to assume $W_i \overset{\perp\!\!\!\perp}{\sim} \mathrm{N}(0, \sigma^2)$

- Assuming that Y follows a normal distribution is equivalent to assuming that T follows a lognormal distribution

- Thus, in the absence of censoring, we could simply use ordinary least squares regression techniques to fit the model, obtain confidence intervals, etc.

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

The Lognormal AFT (cont'd)

- Of course, we will almost always have some censoring, so we will have to extend ordinary linear model methods to handle censoring

- Furthermore, the lognormal distribution, while convenient, does not accurately describe most time-to-event distributions:

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

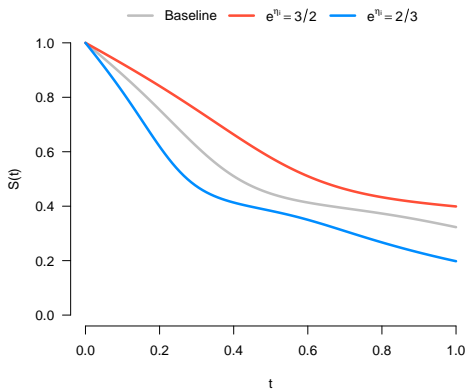## AFTs rescale time

- For any AFT, we have

$$T = e^{\eta} T_0,$$

where $T_0 = e^W$ and $\eta = \mathbf{x}^T \boldsymbol{\beta}$

- In other words, whereas in a proportional hazards (PH) model, the covariates act multiplicatively on the hazard, in an AFT model the covariates act multiplicatively on time:
  - If $e^{\eta_i} = 1/2$, that subject effectively ages at twice normal speed
  - If $e^{\eta_i} = 2$, that subject effectively ages at only half of normal speed
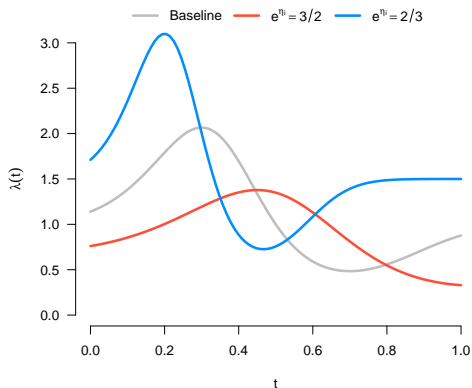
The AFT model framework
Estimation and inference
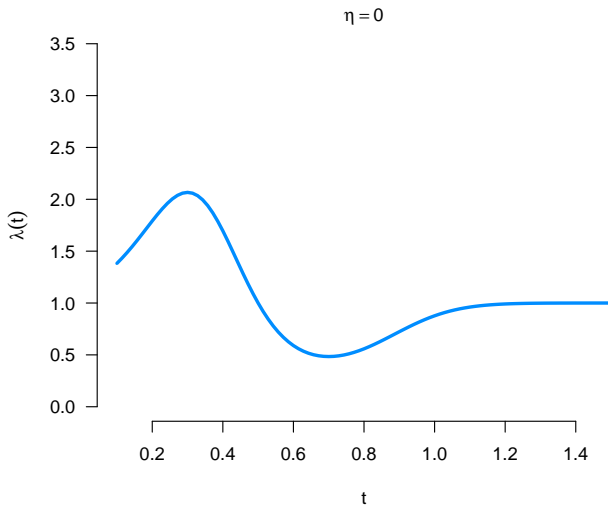survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

Survival

$$S_i(t) = S_0(e^{-\eta_i}t)$$

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

Hazard

$$\lambda_i(t) = \lambda_0(e^{-\eta_i} t) e^{-\eta_i}$$

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

AFT hazard animation

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

PH hazard animation

The AFT model framework
Estimation and inference
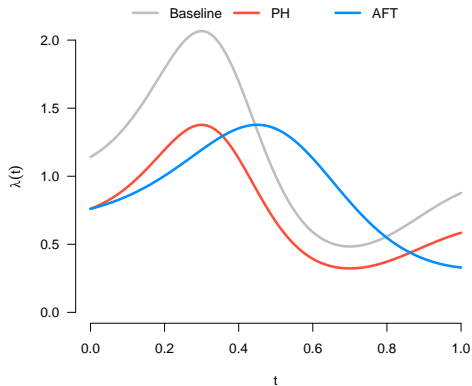survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

AFT vs PH: Hazard ($e^{\eta_i} = 2/3$ for PH; $e^{\eta_i} = 3/2$ for AFT)

$$\text{PH} : \lambda_i(t) = \lambda_0(t)e^{\eta_i}$$
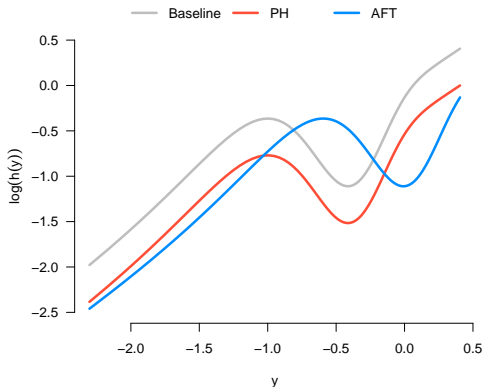$$\text{AFT} : \lambda_i(t) = \lambda_0(e^{-\eta_i}t)e^{-\eta_i}$$

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

AFT vs PH: Log scale ($e^{\eta_i} = 2/3$ for PH, $3/2$ for AFT)

Letting $h(y)$ denote the hazard of y and $f_0(y) = \log \lambda_0(e^y) + y$,

$$\text{PH} : \log h_i(y) = f_0(y) + \eta$$
$$\text{AFT} : \log h_i(y) = f_0(y - \eta)$$

The AFT model framework
Estimation and inference
survreg

Introduction
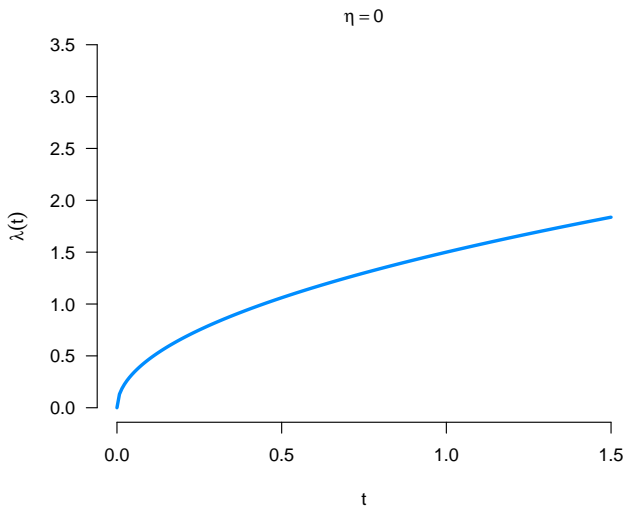Example: The lognormal AFT
Meaning of AFT models

Remarks

- As the final plot depicts, on the log-time vs. log-hazard scale, the effect of the PH assumption is to model the change in hazard as a vertical shift, while the effect of the AFT assumption is to model the change in hazard as a horizontal shift

- In general, the two assumptions cannot be reconciled – on the previous plot, there is no way to shift the gray line vertically and obtain the blue line, for example – and therefore, a model is typically either an AFT model or a PH model, but not both

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

Weibull regression satisfies both AFT and PH

- There is one exception, however: if the distribution is linear (on the log-time vs. log-hazard scale), then any vertical shift of the line will correspond to a horizontal shift

- Recall that the extreme value distribution, $\lambda(y) = e^y$, is linear on this scale, and that the Weibull distribution represents its location-scale family (i.e., all possible intercepts and slopes)

- Thus, the Weibull distribution is the unique distribution (along with its special cases, like the exponential distribution) that satisfies both the PH and AFT assumptions (see section 2.3.4 of our book for a more formal proof)

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

Weibull AFT hazard animation

The AFT model framework
Estimation and inference
survreg

Introduction
Example: The lognormal AFT
Meaning of AFT models

Interpretation of coefficients

- Finally, let us briefly consider the interpretation of the coefficients in an AFT model

- As we have seen, the effect of a δ_j-unit change in covariate j is to multiply the failure time by $\exp(\delta_j \beta_j)$

- For example, if x_j was a treatment indicator and $\beta_j = 0.4$, we could say that patients who received the treatment survived 50% longer than patients who did not receive the treatment ($e^{0.4} = 1.5$)

The AFT model framework
Estimation and inference
survreg

Estimation
Inference

Maximum likelihood estimation: Setup

- Suppose that we can write our AFT model

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \sigma W_i;$$

  note, for example, that the Weibull and lognormal models can be written this way

- Then the likelihood is

$$L(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{d}) = \prod_i \{\sigma^{-1} f(w_i)\}^{d_i} \{S(w_i)\}^{1-d_i}$$

$$= \prod_i \{\sigma^{-1} \lambda(w_i)\}^{d_i} S(w_i),$$

  where f, λ, and S represent the density, hazard, and survival functions for the error distribution, and $w_i = (y_i - \mathbf{x}_i^T \boldsymbol{\beta})/\sigma$

The AFT model framework
Estimation and inference
survreg

Estimation
Inference

Score function(s): Setup

- Note that the likelihood on the previous slide can be written as:

$$\ell(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{d}) = -\sum_i d_i \log \sigma + g(\mathbf{w}),$$

where

$$g(\mathbf{w}) = \sum_i \{d_i \log \lambda(w_i) + \log S(w_i)\}$$

- In what follows, it will be useful to define the vector $\mathbf{a}$, with elements $a_i = \partial g / \partial w_i$

- The motivation for this is that the distribution we choose for W will change g (and therefore $\mathbf{a}$), but otherwise leave the structure of the problem unaffected

The AFT model framework
Estimation and inference
survreg

Estimation
Inference

Score function(s)

- For the most part, estimation proceeds as it did in the exponential regression example we discussed previously, except that we now have to estimate a scale parameter σ as well

- So, for example,

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = -\sigma^{-1} \mathbf{X}^T \mathbf{a}$$

$$\frac{\partial \ell}{\partial \sigma} = -\sigma^{-1}(d + \mathbf{w}^T \mathbf{a}),$$

where $d = \sum_i d_i$

- The above results work for any distribution; only the value of $\mathbf{a}$ changes (as homework, you are asked to derive $\mathbf{a}$ for the extreme value distribution)

The AFT model framework
Estimation and inference
survreg

Estimation
Inference

Newton-Raphson algorithm

- Generally speaking, $\mathbf{a}$ will be a nonlinear function of $\boldsymbol{\beta}$, meaning that to solve $\mathbf{u}(\boldsymbol{\beta}, \sigma) = \mathbf{0}$ we again must use the Newton-Raphson approach we discussed previously

- One complication here is that it is possible to obtain unreasonable updates for σ (e.g., negative values)

- There are various ways to deal with this issue; one is to update σ more gradually by

$$\sigma_{m+1} = (1 - \tau)\sigma_m + \tau\tilde{\sigma},$$

where $\tilde{\sigma}$ is the Newton-Raphson update

The AFT model framework
Estimation and inference
survreg

Estimation
Inference

Information matrix

- Again, the Wald approach is the most convenient, although not the most accurate, for inference
- To proceed, we need (the inverse of) the information matrix:

$$\mathbf{I}(\hat{\theta}) = - \begin{bmatrix} \frac{\partial^2 \ell}{\partial \boldsymbol{\beta}^2} & \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \sigma} \\ \frac{\partial^2 \ell}{\partial \sigma \partial \boldsymbol{\beta}} & \frac{\partial^2 \ell}{\partial \sigma^2} \end{bmatrix};$$

  expressions for these quantities are given in the book
- It is worth noting that the Wald approach is typically more accurate if the likelihood is parameterized in terms of $\tau = \log \sigma$; this is the approach taken in the survival package (this also helps with computation, as negative values are no longer problematic)

The AFT model framework
Estimation and inference
survreg

Estimation
Inference

Accounting for uncertainty in σ

- As we have remarked many times, it is important to note that

$$\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}^2}\right)^{-1}$$

  underestimates the uncertainty with respect to $\boldsymbol{\beta}$

- Instead, we must consider the upper left $p \times p$ block of $\mathbf{I}^{-1}$, which takes into account the fact that uncertainty about σ increases standard errors for $\boldsymbol{\beta}$ (and vice versa)

The AFT model framework
Estimation and inference
survreg

Basic usage

- The `survival` package offers a function `survreg`, for fitting parametric AFT models
- The syntax is similar to other regression modeling functions in R:

```
survreg(S ~ trt + stage + hepato + bili, pbc)
```

where `S` is a `Surv` object

- The default is to use a Weibull distribution, but exponential, lognormal, and other distributions are available using the `dist=` option

The AFT model framework
Estimation and inference
survreg

Exponential regression

- Thus, for example,

```
survreg(S ~ trt + stage + hepato + bili, pbc,
        dist='exponential')
```

- Note that coef(fit) exactly matches our do-it-yourself regression results from last week, except that all the signs are reversed (increasing hazard means decreasing failure time)

- Furthermore, vcov(fit) exactly matches the inverse of our information matrix from last week

The AFT model framework
Estimation and inference
survreg

Weibull results

```
> fit <- survreg(S ~ trt + stage + hepato + bili, pbc)
> summary(fit)

              Value Std. Error       z        p
(Intercept)  9.8086    0.34509  28.423 1.05e-177
trt          0.0960    0.11510   0.834  4.04e-01
stage       -0.4294    0.09045  -4.747  2.06e-06
hepato      -0.2363    0.14625  -1.616  1.06e-01
bili        -0.0915    0.00923  -9.916  3.53e-23
Log(scale)  -0.3760    0.06898  -5.451  5.00e-08

Scale= 0.687
Number of Newton-Raphson Iterations: 6
```

The AFT model framework
Estimation and inference
survreg

Remarks

- Broadly speaking, these results are similar to what we obtained with exponential regression (bilirubin and stage clearly significant, treatment and hepatomegaly not significant)

- The coefficients are somewhat different, of course, and have a different interpretation: a one-unit change in stage shortens survival time by 35% ($e^{-0.4294} = 0.65$)

- Also, note that the summary provides a test of whether $\log \sigma = 0$; in other words, for the adequacy of exponential regression, which is firmly rejected here despite the diagnostic plot looking OK

The AFT model framework
Estimation and inference
survreg

Weibull proportional hazards model

- The survival package does not offer a function for fitting (parametric) proportional hazards models

- Recall, however, that exponential and Weibull proportional hazards models can be reparameterized as AFT models

- Thus, we can obtain the PH coefficients by fitting the AFT model and using the transformation $\boldsymbol{\beta} = -\boldsymbol{\beta}^*/\sigma$, where $\boldsymbol{\beta}$ denotes the PH coefficients and $\boldsymbol{\beta}^*$ denotes the AFT coefficients:

```
b.ph <- -coef(fit.aft)/fit.aft$scale
```

The AFT model framework
Estimation and inference
survreg

Comparison of estimates

	AFT		PH	
	Exponential	Weibull	Exponential	Weibull
trt	0.14	0.10	-0.14	-0.14
stage	-0.56	-0.43	0.56	0.63
hepato	-0.34	-0.24	0.34	0.34
bili	-0.11	-0.09	0.11	0.13

The AFT model framework
Estimation and inference
`survreg`

Predictions

Predicted survival curves for a patient on the placebo arm, stage 2/3/4 cirrhosis, no hepatomegaly, and bili=1