Flexible parametric alternatives to the Cox model, and more

Patrick Royston
MRC Clinical Trials Unit, London

UK Stata User Group, 14 May 2001

Motivation

"The success of Cox regression has perhaps had the unintended side-effect that practitioners too seldomly invest efforts in studying the *baseline hazard* ...

... A *parametric version* [of the Cox model], ... if found to be adequate, would lead to more precise estimation of survival probabilities and ... concurrently contribute to a *better understanding of the phenomenon* under study." (Hjort 1992)

- Smooth the survival function

- Smooth the hazard function

- Models with non-proportional hazards

- Completely specified probability models (e.g. for simulation, validation)

Models based on transformation of $S(t;\mathbf{z})$

Survival function $S(t;\mathbf{z})$ with covariates $\mathbf{z}$

Baseline survival $S_0(t) = S(t;\mathbf{0})$

Consider transformation class of models

$$\boxed{g_\theta[S(t;\mathbf{z})] = g_\theta[S_0(t)] + \boldsymbol{\beta}^\mathsf{T}\mathbf{z}}$$

with

$$g_\theta(S) = \ln\left(\frac{S^{-\theta}-1}{\theta}\right)$$

(Younes & Lachin 1997, Aranda-Ordaz 1981).
Then

$$\begin{aligned} g_0(S) &= \lim_{\theta\to 0}\ln\left(\frac{S^{-\theta}-1}{\theta}\right) = \ln(-\ln S) \\ g_1(S) &= \ln\left(S^{-1}-1\right) = \ln\left(\frac{1-S}{S}\right) \end{aligned}$$

are important special cases.

Won't consider other values of θ

Special case $\theta = 0$: proportional hazards

$$
\begin{aligned}
g_0\left[S(t;\mathbf{z})\right] &= g_0\left[S_0(t)\right] + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{z} \\
\ln\left[-\ln S(t;\mathbf{z})\right] &= \ln\left[-\ln S_0(t)\right] + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{z} \\
\ln H(t;\mathbf{z}) &= \ln H_0(t) + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{z} \\
H(t;\mathbf{z}) &= H_0(t)\exp\left(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{z}\right) \\
h(t;\mathbf{z}) &= h_0(t)\exp\left(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{z}\right).
\end{aligned}
$$

Special case $\theta = 1$: proportional odds (of failure)

$$
\begin{aligned}
g_1\left[S(t;\mathbf{z})\right] &= g_1\left[S_0(t)\right] + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{z} \\
\ln\frac{1 - S(t;\mathbf{z})}{S(t;\mathbf{z})} &= \ln\frac{1 - S_0(t)}{S_0(t)} + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{z} \\
\ln O(t;\mathbf{z}) &= \ln O_0(t) + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{z} \\
O(t;\mathbf{z}) &= O_0(t)\exp\left(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{z}\right).
\end{aligned}
$$

Connexion with binary data models:

PH model has cloglog link
PO model has logistic link

Must approximate baseline distribution function
$H_0(t)$ or $O_0(t)$.

Spline-smoothing baseline distribution functions

Smooth $\ln H_0(t)$ and $\ln O_0(t)$ as fns of $x = \ln t$

Use ‘natural’ cubic regression spline $s(x)$ (constrained to be linear beyond boundary knots)

$$s(x) = \gamma_0 + \gamma_1 x + \gamma_2 v_1(x) + \ldots + \gamma_{m+1} v_m(x)$$

Choose internal knot positions as percentiles of distribution of x for uncensored observations:

When have ‘0 knots’ $s(x)$ is linear—model reverts to Weibull ($\theta = 0$, PH) or log-logistic ($\theta = 1$, PO)

Estimation by ML: `stpm.ado`