Weibull-Cox proportional hazard model

James Barrett
Institute of Mathematical and Molecular Biomedicine,
King's College London

21 July 2014

Abstract

This document contains the mathematical theory behind the Weibull-Cox Matlab function (also called the Weibull proportional hazards model). The likelihood function and it's partial derivatives are given. The Weibull-Cox model assumes a traditional Cox proportional hazards hazard rate but with a Weibull base hazard rate (instead of Breslow's estimator which is implicitly assumed in most implementations of the Cox model). The standard Breslow-Cox model is ill suited to predicting the event time for new individuals. For this purpose the Weibull-Cox model can provide predictions with error bars (given by the standard deviation) along with the usual regression coefficients, survival functions and hazard rates. Requires Optimisation toolbox. Built on Matlab 8.1.0.604 (R2013a).

Background

The Breslow-Cox model is arguably the most popular method of analysing survival data. Breslow's non-parametric estimator of the baseline hazard rate is highly flexible at capturing the time effects in the hazard rate. However it not a smooth estimator (hence the steps in survival curves). It also results in an event time probability density that is not correctly normalised and which consequently cannot be used to make predictions (given by the expected time). A Weibull hazard rate solves both of these problems.

To see this explicitly consider Breslow's estimate of the cumulative hazard rate

$$\hat{\Lambda}_0(\tau) = \sum_{\tau_i \leq \tau} \frac{1}{\sum_{j \in R(\tau_i)} e^{\hat{\boldsymbol{\beta}} \cdot \mathbf{x}_j}} \tag{1}$$

which and was originally presented in the discussion section of **?**. The regression parameters $\hat{\boldsymbol{\beta}}$ are maximum likelihood estimators obtained from the partial likelihood. We assume we have observed survival data $D = \{(\tau_1, \Delta_1), \ldots, (\tau_N, \Delta_N)\}$ for $i = 1, \ldots, N$ individuals where $\tau_i > 0$ is the time until an event and $\Delta_i = 0$ indicates right censoring (assumed independent) and $\Delta_i = 1$ indicates the primary event occurred for individual i.

Once $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0(\tau)$ have been inferred from observed data the event time density corresponding to an individual with covariates $\mathbf{x}^*$ is

$$p(\tau|\mathbf{x}^*, \hat{\boldsymbol{\beta}}, \hat{\Lambda}_0) = \lambda_0(\tau_i) e^{\boldsymbol{\beta} \cdot \mathbf{x}^*} \exp(-\Lambda_0(\tau) e^{\boldsymbol{\beta} \cdot \mathbf{x}^*}). \tag{2}$$

However, the event time density is not normalised. To see this consider

$$\int_0^\infty \mathrm{d}s\, p(s|\mathbf{x}^*, \hat{\boldsymbol{\beta}}, \hat{\Lambda}_0) = 1 - \exp(-\hat{\Lambda}_0(s)e^{\hat{\boldsymbol{\beta}}\cdot\mathbf{x}^*})\Big|_{s=\infty}. \tag{3}$$

Correct normalisation requires

$$\lim_{\tau\to\infty} \Lambda(\tau) = \infty, \tag{4}$$

a condition that is not met by Breslow's estimator since the largest value (**??**) can take occurs after the largest observed time $\max_i(\tau_i)$ and which we denote with $\hat{\Lambda}_0^\infty$

$$\hat{\Lambda}_0^\infty = \sum_{\tau_i} \frac{1}{\sum_{j\in R(\tau_i)} e^{\hat{\boldsymbol{\beta}}\cdot\mathbf{x}_j}} < \infty. \tag{5}$$

Nevertheless survival curves can be generated according to

$$S(\tau|\mathbf{x}^*, \hat{\boldsymbol{\beta}}, \hat{\Lambda}_0) = \exp(-\hat{\Lambda}_0(\tau)e^{\hat{\boldsymbol{\beta}}\cdot\mathbf{x}^*}). \tag{6}$$

Weibull-Cox model definition

We choose a Weibull base hazard rate

$$\lambda_0(\tau) = (\nu/\rho)(\tau/\rho)^{\nu-1} \tag{7}$$

where $\rho > 0$ is a scale parameter and $\nu > 0$ is a shape parameter. It follows that the cumulative base hazard rate is $\Lambda_0(\tau) = (\tau/\rho)^\nu$. Note that the normalisation condition (**??**) is satisfied. The hazard rate for individual i is

$$\pi_i(\tau|\mathbf{x}_i, \nu, \rho, \boldsymbol{\beta}) = \lambda_0(\tau)e^{\boldsymbol{\beta}\cdot\mathbf{x}_i}, \tag{8}$$

where $\mathbf{x}_i \in \mathbb{R}^q$ is a vector of covariates. Using Bayes' theorem the posterior over parameters is $p(\boldsymbol{\beta}, \rho, \nu|D) \propto p(D|\boldsymbol{\beta}, \rho, \nu)p(\boldsymbol{\beta})p(\rho)p(\nu)$. The data likelihood is given by

$$p(D|\boldsymbol{\beta}, \lambda_0) = \prod_{i=1}^N [\lambda_0(\tau_i)e^{\boldsymbol{\beta}\cdot\mathbf{x}_i}]^{\Delta_i} \exp(-\Lambda_0(\tau)e^{\boldsymbol{\beta}\cdot\mathbf{x}_i}). \tag{9}$$

We can then define the log likelihood as

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, \rho, \nu) &= -\frac{1}{N}\log p(\boldsymbol{\beta}, \rho, \nu|D) \\ &= -\frac{1}{N}\sum_{i:\Delta_i=1}[\log\lambda_0(\tau_i) + \boldsymbol{\beta}\cdot\mathbf{x}_i] + \frac{1}{N}\sum_{i=1}^N \Lambda_0(\tau_i)e^{\boldsymbol{\beta}\cdot\mathbf{x}_i} \\ &\quad - \frac{1}{N}\log p(\boldsymbol{\beta}) - \frac{1}{N}\log p(\rho) - \frac{1}{N}\log p(\nu)\end{aligned} \tag{10}$$

where $\log\lambda_0(\tau) = \log\nu - \log\rho + (\nu-1)\log(\tau/\rho)$. We assume $p(\boldsymbol{\beta})$, $p(\rho)$ and $p(\nu)$ are constant (and therefore improper) priors. The optimal values of the parameters are given by numerically solving

$$\{\hat{\boldsymbol{\beta}}, \hat{\rho}, \hat{\nu}\} = \mathrm{argmax}_{\boldsymbol{\beta},\rho,\nu}\mathcal{L}(\boldsymbol{\beta}, \rho, \nu). \tag{11}$$

The gradient based `fminunc` optimisation function is used. Partial derivatives can be found below.

Finally, error bars for β_r can be obtained from $\sqrt{(N\mathbf{H})^{-1}_{rr}}$. This gives the standard deviation of that parameter under a Gaussian approximation of the posterior. The matrix $\mathbf{H}$ is defined below. Error bars for ρ and ν are not defined under a Gaussian approximation since by definition both parameters are non-negative. In the implementation both parameters are reparameterised such that they take real values.

Predictions

Predictions can be made by computing the mean (and variance) of the event time density corresponding to a new individual with covariates $\mathbf{x}^*$

$$\langle \tau \rangle = \int_0^\infty \mathrm{d}s\, s\lambda_0(s)e^{\hat{\boldsymbol{\beta}}\cdot\mathbf{x}^*}\exp(-\Lambda_0(s)e^{\hat{\boldsymbol{\beta}}\cdot\mathbf{x}^*}). \tag{12}$$

The hazard rate and survival function are respectively given by

$$\pi(\tau|\mathbf{x}^*, \hat{\boldsymbol{\beta}}, \hat{\rho}, \hat{\nu}) = (\hat{\nu}/\hat{\rho})(\tau/\hat{\rho})^{\hat{\nu}-1}e^{\hat{\boldsymbol{\beta}}\cdot\mathbf{x}^*} \tag{13}$$

$$S(\tau|\mathbf{x}^*, \hat{\boldsymbol{\beta}}, \hat{\rho}, \hat{\nu}) = e^{(\tau/\rho)^\nu e^{\hat{\boldsymbol{\beta}}\cdot\mathbf{x}^*}}. \tag{14}$$

Usage

A model is fit with the `model = wc_train(X, t, E)` function where `X` is a matrix of covariates, `t` is a vector of the event times and `E` is a vector of indictor variables. Type `help wc_train` for full details. The function returns a structure which contains the inferred values of the parameters along with error bars (standard deviations).

Once a `model` structure has been trained predictions can be made with the `wc_predict` function. Survival curves and hazard rates can be generated using `wc_survival` and `wc_hazard` respectively. Use `help` for further information. An example is given in `wc_example.m`.

Example

This is a one dimensional dataset with $N = 25$ that were generated synthetically. Results are presented in Figures **??**, **??**, and **??**. We can also compare the results to a standard Breslow-Cox analysis. Note that we do note expect $\boldsymbol{\beta}$ to be the same in both cases. We will load a standard matlab example dataset:

```
cd(matlabroot)
cd('help/toolbox/stats/examples/')
load readmissiontimes
X = [Age, Sex, Weight];
[b,logl,H,stats] = coxphfit(X,ReadmissionTime,'censoring',Censored);
```

We can then run the Weibull-Cox model. Note that `coxphfit` uses different censoring labels.

```
model =  wc_train(X, ReadmissionTime, 1 - Censored);
```

The $\boldsymbol{\beta}$ coefficients with standard deviations are

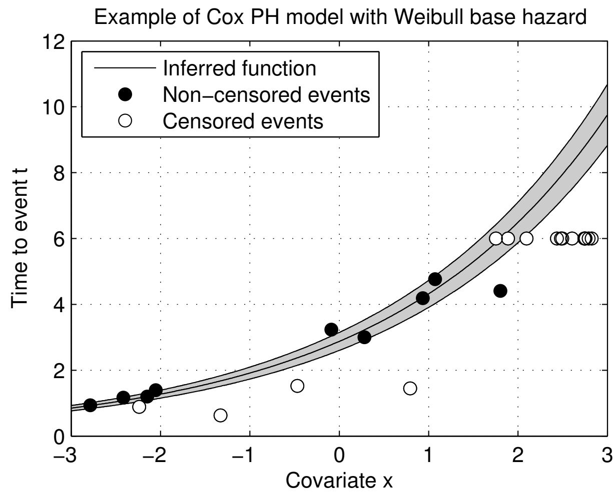

Figure 1: Plot of the observed event times as a function of the covariate x. The white circles represent censoring events. Note that an 'end of trial' cutoff was imposed at 6 years. The black line is the predicted time as a function of the covariate obtained using (**??**). It can be interpreted as an inferred function relating the event times to the covariates. The grey area represents plus and minus one standard deviation.

Breslow-Cox:	beta	std	Weibull-Cox:	beta	std
	0.01	0.02		0.01	0.00
	-0.54	0.80		-1.78	0.04
	0.02	0.02		0.00	0.00

Partial derivatives

We require partial derivatives of the log likelihood for the Weibull-Cox model. These are

$$\frac{\partial}{\partial \beta_s}\mathcal{L}(\boldsymbol{\beta},\rho,\nu) = -\frac{1}{N}\sum_{i:\Delta_i=1} x_{is} + \frac{1}{N}\sum_{i=1}^{N}\Lambda_0(\tau_i)x_{is}e^{\boldsymbol{\beta}\cdot\mathbf{x}_i} \tag{15}$$

and

$$\frac{\partial}{\partial \rho}\mathcal{L}(\boldsymbol{\beta},\rho,\nu) = \frac{N_1}{N}\frac{\nu}{\rho} + \frac{1}{N}\sum_{i=1}^{N}\frac{\partial\Lambda_0(\tau_i)}{\partial\rho}e^{\boldsymbol{\beta}\cdot\mathbf{x}_i} \tag{16}$$

$$\frac{\partial}{\partial \nu}\mathcal{L}(\boldsymbol{\beta},\rho,\nu) = -\frac{N_1}{N}\frac{1}{\nu} - \frac{1}{N}\sum_{i:\Delta_i=1}\log(\tau_i/\rho) + \frac{1}{N}\sum_{i=1}^{N}\frac{\partial\Lambda_0(\tau_i)}{\partial\nu}e^{\boldsymbol{\beta}\cdot\mathbf{x}_i} \tag{17}$$

where we have used

$$\frac{\partial}{\partial\rho}\log\lambda_0(\tau) = -\frac{\nu}{\rho} \tag{18}$$

$$\frac{\partial}{\partial\nu}\log\lambda_0(\tau) = \frac{1}{\nu} + \log(\tau/\rho) \tag{19}$$

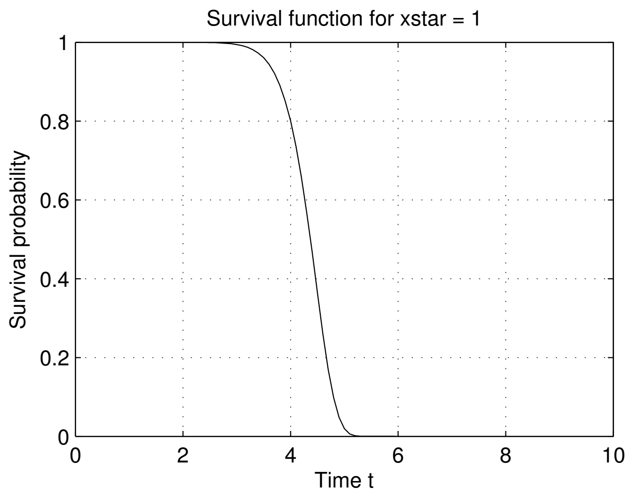

Figure 2: A plot of the survival function for an individual with $x^* = 1$ obtained from (**??**). One can see the survival probability drops rapidly around $t = 5$ which is consistent with Figure **??**.

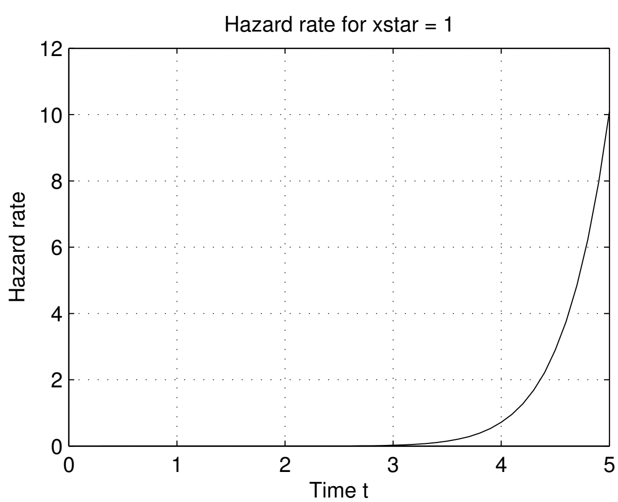

Figure 3: A plot of the hazard function for an individual with $x^* = 1$ obtained from (**??**). One can see the hazard rate increases rapidly around $t = 5$ which is also consistent with Figures **??** and **??**.

and

$$\frac{\partial \Lambda_0(\tau)}{\partial \rho} = -\frac{\nu}{\rho}\left(\frac{\tau}{\rho}\right)^{\nu} \tag{20}$$

$$\frac{\partial \Lambda_0(\tau)}{\partial \nu} = (\log \tau - \log \rho)\left(\frac{\tau}{\rho}\right)^{\nu}. \tag{21}$$

Since we require $\rho > 0$ we write it in the form

$$\rho = (1 + \rho_{LB} + \exp(\tilde{\rho})) \tag{22}$$

where $\tilde{\rho} \in \mathbb{R}$ and $\rho_{LB} \geq 0$ is a lower bound on ρ that can be set manually. This formulation allows the use of unconstrained optimisation functions to be used. However the partial derivatives now become

$$\frac{\partial \mathcal{L}}{\partial \tilde{\rho}} = \frac{\partial \mathcal{L}}{\partial \rho}\frac{\partial \rho}{\partial \tilde{\rho}} \quad \text{with} \quad \frac{\partial \rho}{\partial \tilde{\rho}} = \frac{e^{\tilde{\rho}}}{1 + e^{\tilde{\rho}}}. \tag{23}$$

We also require $\nu > 0$ and the same formulation is used.

Second order partial derivatives

$$\frac{\partial^2}{\partial \beta_r \partial \beta_s}\mathcal{L}(\boldsymbol{\beta}, \rho, \nu) = \frac{1}{N}\sum_{i=1}^{N} \Lambda_0(\tau_i) x_{is} x_{ir} e^{\boldsymbol{\beta}\cdot\mathbf{x}_i} \tag{24}$$

and

$$\frac{\partial^2}{\partial \rho^2}\mathcal{L}(\boldsymbol{\beta}, \rho, \nu) = -\frac{N_1}{N}\frac{\nu}{\rho^2} + \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\nu(\nu+1)}{\rho^2}\left(\frac{\tau_i}{\rho}\right)^{\nu}\right] e^{\boldsymbol{\beta}\cdot\mathbf{x}_i} \tag{25}$$

$$\frac{\partial^2}{\partial \nu^2}\mathcal{L}(\boldsymbol{\beta}, \rho, \nu) = \frac{N_1}{N}\frac{1}{\nu^2} + \frac{1}{N}\sum_{i=1}^{N}(\log \tau_i - \log \rho)^2 \left(\frac{\tau_i}{\rho}\right)^{\nu} e^{\boldsymbol{\beta}\cdot\mathbf{x}_i}. \tag{26}$$

Finally we require

$$\frac{\partial^2}{\partial \nu \partial \rho}\mathcal{L}(\boldsymbol{\beta}, \rho, \nu) = \frac{\partial^2}{\partial \rho \partial \nu}\mathcal{L}(\boldsymbol{\beta}, \rho, \nu) = \frac{N_1}{N}\frac{1}{\rho} - \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\nu}{\rho}(\log \tau_i - \log \rho)\left(\frac{\tau_i}{\rho}\right)^{\nu} + \frac{1}{\rho}\left(\frac{\tau_i}{\rho}\right)^{\nu}\right] \tag{27}$$

$$\frac{\partial^2}{\partial \rho \partial \beta_s}\mathcal{L}(\boldsymbol{\beta}, \rho, \nu) = -\frac{1}{N}\frac{\nu}{\rho}\sum_{i=1}^{N} x_{is}\left(\frac{\tau_i}{\rho}\right)^{\nu} e^{\boldsymbol{\beta}\cdot\mathbf{x}_i} \tag{28}$$

$$\frac{\partial^2}{\partial \nu \partial \beta_s}\mathcal{L}(\boldsymbol{\beta}, \rho, \nu) = \frac{1}{N}\sum_{i=1}^{N}(\log \tau_i - \log \rho) x_{is}\left(\frac{\tau_i}{\rho}\right)^{\nu} e^{\boldsymbol{\beta}\cdot\mathbf{x}_i} \tag{29}$$

Since we have written the parameters ρ and ν in the form (**??**) the second order partial derivatives are in practice given by

$$\frac{\partial^2 \mathcal{L}}{\partial \tilde{\rho}^2} = \frac{\partial^2 \mathcal{L}}{\partial \rho^2} \left(\frac{\partial \rho}{\partial \tilde{\rho}} \right)^2 + \frac{\partial \mathcal{L}}{\partial \rho} \frac{\partial^2 \rho}{\partial \tilde{\rho}^2} \quad \text{with} \quad \frac{\partial^2 \rho}{\partial \tilde{\rho}^2} = \frac{e^{\tilde{\rho}}}{(1 + e^{\tilde{\rho}})^2}, \tag{30}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \tilde{\rho} \partial \tilde{\nu}} = \frac{\partial^2 \mathcal{L}}{\partial \tilde{\nu} \partial \tilde{\rho}} = \frac{\partial^2 \mathcal{L}}{\partial \rho \partial \nu} \frac{\partial \rho}{\partial \tilde{\rho}} \frac{\partial \nu}{\partial \tilde{\nu}}, \tag{31}$$

$$\frac{\partial^2 \mathcal{L}}{\partial \tilde{\rho} \partial \beta_s} = \frac{\partial^2 \mathcal{L}}{\partial \rho \partial \beta_s} \frac{\partial \rho}{\partial \tilde{\rho}} \text{ and } \frac{\partial^2 \mathcal{L}}{\partial \tilde{\nu} \partial \beta_s} = \frac{\partial^2 \mathcal{L}}{\partial \nu \partial \beta_s} \frac{\partial \nu}{\partial \tilde{\nu}}. \tag{32}$$