Weibull survivals with changepoints and heterogeneity

Ana-María Lara-Porras, Esteban Navarrete-Álvarez, Julia García-Leal, and José-Manuel Quesada-Rubio

Department of Statistics and Operations Research, Faculty of Sciences,
University of Granada,
Campus de Fuentenueva, s/n, 18071, Granada, Spain
(e-mail: alara@ugr.es)

Abstract. This paper describes the proportional hazard models of right-censored survival data with Weibull distribution whose parameters vary in the time and the impact of individual heterogeneity being described by frailty. We obtain the equations of the maximum likelihood estimators for this model.
Keywords: proportional hazard models, Weibull distribution, frailty.

1 Introduction

The proportional hazard model is found among the most important models for the survival analysis data. The use of proportional hazard models for the study of survival times has received a great attention on the part of researchers, both in their theoretical aspect as well as in their interesting and numerous applications.

Likewise, an important characteristic of this type of investigations is that the data are frequently incomplete, meaning that the observation of survival time is not known for all individuals. These data are known as censored data.

[Aitkin and Clayton, 1980] and [Noura and Read, 1990] study the completely parametric models with a specified baseline hazard distribution. The first authors make use of exponential, Weibull, extreme value and generalized extreme value distributions. The second authors consider the piecewise model of the baseline hazard distribution and study the case of the Weibull distribution. In both works the presence of censored and uncensored observations is considered.

Piecewise models are based on the assumption that the parameters that characterize the base distribution vary with the passing of time. This is the reason why a partition of the time interval is introduced so as to maintain the base distribution, although with different parameters on each one of the intervals.

We have used this method considering that the survival time follows a Weibull distribution where the parameters chraracterizing the base-line distribution may vary with time for different intervals but remain constant at each interval. The points where the parameters change are called "changepoints".

Ordinary life table analyses implicitly assume that the population is homogeneous, an assumption which is usually unrealistic. It is more relevant to consider the population as a mixture of individuals with different risk, the heterogeneity being described by a quantity know as the frailty. Models for heterogeneity have been proposed, for example by [Vaupel *et al*, 1979] who introduced an unobserved quantity, that is the so-called frailty. This quantity described risk factors measurable or nonmeasurable, not included in the model.

The model to describe the population as a mixture assumes that to each individual corresponds a quantity, the frailty, describing the individual's relative risk.

In this work we consider the piecewise model for the Weibull distribution, the existence of censored observations and also we study the heterogeneity between individuals by a fraitly that we suppose follow a positive stable distribution.

2 Model construction

Let the survival time T be a nonnegative random variable that follows Weibull's distribution with a survivor function $S(t)$ and a hazard function $h(t)$. The heterogeneousness of the population is stated by a covariates vector $z = (z_1, z_2, \cdots, z_p)^T$ describing the characteristics of both the patient and the illness.

The hazard function depends in general on both time and on the set of covariates. The proportional hazard model separates these components as,

$$h(t; \mathbf{z}) = \gamma\rho(\rho t)^{\gamma-1} e^{\beta^T \mathbf{z}},$$

where the linear predictor $\beta^T z$ expresses the relative effect of the covariates z in terms of an unknown parameter vector $\beta = (\beta_1, \beta_2, \cdots, \beta_p)^T$.

The survivor function for these models is:

$$S(t; \mathbf{z}) = \exp\left[-(\rho t)^\gamma \exp\left(\beta^T \mathbf{z}\right)\right].$$

The hazard at time t conditional on x for a person with frailty x is assumed to be of form

$$h(t, x) = xh(t),$$

where the non-random function $h(t)$ common for all individuals is independent of x and describes the time effect.

Several authors have studied the model with gamma distributed frailties. We consider a stable positive distribution and whose scale factors have Laplace transform

$$L(s) = \mathrm{E}\left[\exp\left(-sx\right)\right] = \exp\left(-s^\alpha\right) \quad , \quad s \geq 0,$$

where $\alpha \in (0,1]$

Given the Fraitly x these expressions become

$$\begin{aligned} h(t;\mathbf{z},x) &= \gamma\rho(\rho t)^{\gamma-1} x e^{\beta^T \mathbf{z}} \quad , \\ S(t;\mathbf{z},x) &= \exp\left[-(\rho t)^{\gamma} x \exp\left(\beta^T \mathbf{z}\right)\right] . \end{aligned} \tag{1}$$

In this case, the corresponding survivor function is

$$S(t;\mathbf{z}) = \int \exp\left[-\Lambda(t) x e^{\beta^T \mathbf{z}}\right] f(x) dx, \tag{2}$$

where $\Lambda(t)$ is the cumulative hazard function.

If x follows a positive stable function, you get

$$S(t;\mathbf{z}) = \exp\left[-\left[\exp\left(\beta^T \mathbf{z}\right) \Lambda(t)\right]^{\alpha}\right] \tag{3}$$

where α is a parameter coming from the fraitly distribution.

We consider that the parameters that characterize the Weibull distribution can vary with time. Thus we divide the time axis in $k+1$ intervals, by using the changepoints $a_1, \cdots, a_k$. For convenience $a_0 = 0$ and $a_{k+1} = \infty$. In each interval (a_{j-1}, a_j) the distribution parameters take the values ρ_j and γ_j.

Denoting $g(t) = \ln \Lambda(t)$. In $a_{j-1} < t \leq a_j$, $g(t)$ becomes

$$g(t) = \ln\left[\rho_j t\right]^{\gamma_j} \quad , \quad j = 1, \ldots, k+1. \tag{4}$$

For $j = 1, \ldots, k$, due the continuity of $g(t)$ in the changepoints, has to be verified

$$\ln\left[\rho_j a_j\right]^{\gamma_j} = \ln\left[\rho_{j+1} a_j\right]^{\gamma_{j+1}} \quad , \quad j = 1, \ldots, k, \tag{5}$$

from where we derive that

$$\gamma_j = \gamma_1 \prod_{p=1}^{j-1} \frac{\ln\left[\rho_p a_p\right]}{\ln\left[\rho_{p+1} a_p\right]} \quad , \quad j = 2, \ldots, k+1. \tag{6}$$

Thus, for a survival time ending at $j-$th interval,

$$g(t) = \gamma_1 \ln\left(\rho_j t\right) \prod_{p=1}^{j-1} \frac{\ln\left[\rho_p a_p\right]}{\ln\left[\rho_{p+1} a_p\right]}. \tag{7}$$

For the $i-$th individual

$$g(t_i) = \sum_{j=1}^{k+1} c_{ij} \gamma_1 \ln\left(\rho_j t_i\right) \prod_{p=1}^{j-1} \frac{\ln\left[\rho_p a_p\right]}{\ln\left[\rho_{p+1} a_p\right]}, \tag{8}$$

where for $j = 1$, the product in p is omitted and c_{ij} is an indicator variable defined by:

$$c_{ij} = \begin{cases} 1 \text{ if } a_{j-1} < t_i \leq a_j \\ 0 \text{ otherwise} \end{cases}$$

with $i = 1, \ldots, N$ and $j = 1, \ldots, k+1$ where N represents the number of individuals.

Let $H_i = \exp \alpha g(t_i) + \beta^T \mathbf{z}$ and

$$h_i = H_i' = \alpha g'(t_i) H_i = \alpha H_i \prod_{j=1}^{k+1} \left[\frac{\gamma_1}{t_i} \prod_{p=1}^{j-1} \frac{\ln [\rho_p a_p]}{\ln [\rho_{p+1} a_p]} \right]^{c_{ij}}, \tag{9}$$

the survival and density functions can be expressed by

$$S(t_i; \mathbf{z}) = \exp[-H_i] \quad ; \quad f(t_i; \mathbf{z}) = h_i \exp[-H_i]. \tag{10}$$

3 Likelihood equations

Suppose that in a data set consisting of N observations, n are uncensored and m are censored. We define a censor indicator in the following manner:,

$$\omega_i = \begin{cases} 1 \text{ if the observation is uncensored} \quad (T_i = t_i) \\ 0 \text{ if it is censored} \quad (T_i > t_i) \end{cases}. \tag{11}$$

If a survival time observation is no censored contributes with $f(t)$ to the likelihood and if the observation is censored in time t, contributes with $S(t)$. Thus the likelihood function is,

$$l = \prod_{i=1}^{N} [f(t_i; \mathbf{z})]^{\omega_i} [S(t_i; \mathbf{z})]^{1-\omega_i}, \tag{12}$$

and the log-likelihood function,

$$L = \sum_{i=1}^{N} \{\omega_i \ln h(t_i; \mathbf{z}) + \ln S(t_i; \mathbf{z})\} =$$

$$\sum_{i=1}^{N} \left\{ \omega_i \left[\ln \alpha + \alpha \left(\sum_{s=1}^{p} \beta_s z_{is} + \sum_{j=1}^{k+1} c_{ij} \left(\gamma_1 \ln (\rho_j t_i) \prod_{p=1}^{j-1} \frac{\ln [\rho_p a_p]}{\ln [\rho_{p+1} a_p]} \right) \right) + \right. \right.$$

$$\left. \sum_{j=1}^{k+1} c_{ij} \ln \left(\frac{\gamma_1}{t_i} \prod_{p=1}^{j-1} \frac{\ln [\rho_p a_p]}{\ln [\rho_{p+1} a_p]} \right) \right] -$$

$$\left. \exp \alpha \left[\sum_{s=1}^{p} \beta_s z_{is} + \sum_{j=1}^{k+1} c_{ij} \left(\gamma_1 \ln (\rho_j t_i) \prod_{p=1}^{j-1} \frac{\ln [\rho_p a_p]}{\ln [\rho_{p+1} a_p]} \right) \right] \right\}. \tag{13}$$

where for $j = 1$, the product in p is omitted.

The first derivatives of L with respect to the parameters β_s, γ, α and ρ are **:**

$$\frac{\partial L}{\partial \beta_s} = \alpha \sum_{i=1}^{N} z_{is} \left(\omega_i - H_i\right) \qquad \text{para} \qquad s = 1, \ldots, p \tag{14}$$

$$\frac{\partial L}{\partial \gamma_1} = \sum_{i=1}^{N} \alpha \left\{ \left(\omega_i - H_i\right) \left[\sum_{j=1}^{k+1} c_{ij} \ln\left(\rho_j t_i\right) \prod_{p=1}^{j-1} \frac{\ln\left[\rho_p a_p\right]}{\ln\left[\rho_{p+1} a_p\right]} \right] + \frac{\omega_i}{\gamma_1} \right\} \tag{15}$$

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{N} \left\{ \frac{\omega_i}{\alpha} + \left(\omega_i - H_i\right) \left[\sum_{s=1}^{p} \beta_s z_{is} + \sum_{j=1}^{k+1} c_{ij} \left(\gamma_1 \ln\left(\rho_j t_i\right) \prod_{p=1}^{j-1} \frac{\ln\left[\rho_p a_p\right]}{\ln\left[\rho_{p+1} a_p\right]} \right) \right] \right\} \tag{16}$$

$$\frac{\partial L}{\partial \rho_1} = \frac{1}{\rho_1} \sum_{i=1}^{N} \left\{ \alpha\gamma_1 \left(\omega_i - H_i\right) \left(c_{i1} + \sum_{j=2}^{k+1} c_{ij} \frac{\ln\left(\rho_j t_i\right)}{\ln\left(\rho_2 a_1\right)} \prod_{p=2}^{j-1} \frac{\ln\left[\rho_p a_p\right]}{\ln\left[\rho_{p+1} a_p\right]} \right) + \frac{\omega_i}{\ln\left(\rho_1 a_1\right)} \sum_{j=2}^{k+1} c_{ij} \right\} \tag{17}$$

$$\frac{\partial L}{\partial \rho_j} = \frac{1}{\rho_j} \sum_{i=1}^{N} \left\{ \left(\omega_i - H_i\right) \frac{\alpha\gamma_1}{\left(\ln\left(\rho_j a_{j-1}\right)\right)^2} \frac{\prod_{p=1}^{j-1} \ln\left[\rho_p a_p\right]}{\prod_{p=2}^{j-1} \ln\left[\rho_p a_{p-1}\right]} \times \left[c_{ij} \ln \frac{a_{j-1}}{t_i} + \frac{\ln \frac{a_{j-1}}{a_j}}{\ln\left[\rho_{j+1} a_j\right]} \sum_{p=j+1}^{k+1} c_{ip} \ln\left(\rho_p t_i\right) \prod_{r=j+1}^{p-1} \frac{\ln\left[\rho_r a_r\right]}{\ln\left[\rho_{r+1} a_r\right]} \right] + \frac{\omega_i}{\ln\left(\rho_j a_{j-1}\right)} \left(-c_{ij} + \frac{\ln \frac{a_{j-1}}{a_j}}{\ln\left[\rho_j a_j\right]} \sum_{p=j+1}^{k+1} c_{ip} \right) \right\} \qquad \text{for} \qquad j = 2, \ldots, k+1 \ . \tag{18}$$

4 Some questions about the equations resolution

The solving of these equations may be performed by general iterative methods, by directly employing statistical packages such as GLIM or we can study the behaviour of maximum-likelihood estimators through the simulation.

Upon the determination of the appropriate number of changepoints and their locations there are some graphic procedures. In practice it is enough, in most cases, to consider only one or two changepoints. Moreover, it must be indicated that the physical nature of the problem also sometimes permits on the location of the possible changepoints.

References

[Aalen, 1988]O. Aalen. Heterogeneity in Survival Analysis. *Statistics in Medicine*, 1121–1137, 1988.

[Aitkin and Clayton, 1980]M. Aitkin and D. Clayton. The Fitting of Exponential, Weibull and Extreme Value Distributions To Complex Censored Survival Data using GLIM. *Applied Statistics*, 156–163, 1980.

[Cox and Oakes, 1984]D.R. Cox and D. Oakes. *Analysis of Survival Data*. Chapman and Hall, 1984.

[Hougaard, 1984]P. Hougaard. Life Table Methods for Heterogeneous Populations: Distributions Describing the Heterogeneity. *Biometrika*, 75–83, 1984.

[Hougaard, 1986]P. Hougaard. Survival Models for Heterogeneous Populations Derived from Stable Distributions. *Biometrika*, 387–96, 1986.

[Hougaard, 1987]P. Hougaard. Modelling Multivariate Survival. *Scand. J. Statist.*, 291–304, 1987.

[Kalbfleisch and Prentice, 1980]J.D. Kalbfleisch and R.L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, New York, 1980.

[Lara-Porras, 1995]A.M. Lara-Porras. *Aportaciones a modelos de supervivencia: distribuciones base con puntos de cambio y covariables dependientes del tiempo*, Unpublished P.H.D. dissertation, University of Granada, Spain, 1995.

[Lara-Porras *et al*, 1998]A.M. Lara-Porras, J. García-Leal, and E. Navarrete-Álvarez. A Proportional Hazard Model with Time-Dependent Parameters and Covariates, *Journal of the Italian Statistical Society*, 233–242, 1998.

[Lara-Porras *et al*, 2000]A.M. Lara-Porras, J. García-Leal, and E. Navarrete-Álvarez. Simulation in a fully parametric proportional hazard model with changepoints, *Brazilian Journal of Probability and Statistics*, 113–122, 2000.

[Noura and Read, 1990]A.A. Noura and K.L.Q. Read. Proportional Hazard Changepoints Models in Survival Analysis, *Appl. Statist.*, 241–253, 1990.

[Vaupel *et al*, 1979]J.W. Vaupel, K.G. Manton, and E. Stallard. The Impact of Heterogeneity in Individual Frailty and the Dynamics of Mortality, *Demography*, 439–454, 1979.