AJAS / SPAS

ISSN 2316-0861

African Journal of Applied Statistics
Vol. 6 (1), 2018, pages 551 - 563.
DOI: http://dx.doi.org/10.16929/ajas/551.230

ISSN 2316-0861

Bayesian Estimation of Cox Proportional Hazard model under a Piecewise Constant Baseline Hazard Function and the problem of Survival Time Axis Grid

Chellai Fatih [1,*], **and Boudrissa Naima** [2]

[1] Department of the Basic Education, University of Ferhat Abbas, Setif, Algeria.
[2] Department of Applied Statistics, High National School of Statistics and Applied Economics, Algeria.

Received on December 12, 2018; Published online March 05, 2019

Abstract. The determination of the number and the lengths of intervals of the baseline risk function $\lambda_0(t)$, is an important issue in Piecewise constant Exponential Models (PEM) and Proportional Hazard Models (PH), especially, when using Bayesian inference. In this context, we propose a simple method to estimate that number and thoses lenghts of interval for constructing Bayesian PH Cox model. Based on real data, the obtained results that the estimated parameters are not affected, but the log-likelihood and information criterion are very sensitive. On this, the problem of model selection is considered to assess the influence on decision making.

Key words: Proportional Harzard Cox model; Survival analysis; Baseline Hazard function; Bayesian methods
AMS 2010 Mathematics Subject Classification Objects : 62F15; 62Nxx.

Presented by Prof. Gane Samb Lo, Gaston Berger University, Saint-Louis (Senegal)
African University od Sciences and Technology, Abuja (Nigeria)
Editor in Chief

* Chellai Fatih: fatih.chellai@univ-setif.dz
Boudrissa Naima : boudrina2005@yahoo.com

Résumé. La détermination du nombre et de la longueur des intervalles de la fonction de risque de base $\lambda_0(t)$, est un problème important dans les modèles exponentiels constant par morceaux (ECM) et les modèles à risque proportionnels (PH); notamment, lorsque nous utilisons l'inférence bayésienne. Dans ce contexte, nous proposons une méthode simple permettant d'estimer le nombre et la longueur des intervalles afin de construire un modèle bayésien de Cox PH. Sur la base des données réelles, les résultats obtenus montrent que les paramètres d'estimation n'ont pas été affectés, mais que les critères de log-vraisemblance et d'information sont très sensible. Sur cette base, le problème de la sélection du modèle est considéré pour évaluer l'influence sur la prise de décision.

1. Introduction

An important issue in any Bayesian analysis is the specification of a prior distribution. This is especially true in survival analysis when one wishes to assess the importance of certain prognostic factors such as age, gender, or a certain treatment in predicting survival outcomes, Seong*et al.* (1986). One of the major problems of the Piecewise-Constant Exponential (PCE) models Friedman (1982), is to determine the number of time intervals (classes) to use. The number of time intervals must be determined by the analyst, although one can choose a certain number of periods. At the level of statistical modeling, it is important to recognize that there is always a compromise to be made. If a large number of periods is chosen, a better approximation of the unknown base risk will be obtained, but a larger number of coefficients will have to be estimated, which may cause problems. Alternatively, if a small number of periods is chosen, there will be fewer estimation problems, but the baseline risk approximation will be worse.

Demarqui*et al.* (2008) assume that the time grid needed to fit the Piece-wise Exponential Model(PME) is a random quantity and propose a flexible class of prior distributions for modeling jointly the time grid and its corresponding failure rates. Although PEM has been widely used in the literature, the time grid $t = t_0, t_1, ..., t_j$ has been chosen arbitrarily in most of these works. In their book, Kalbfleisch and Prentice (2011) suggest that the selection of the time grid $t = t_0, t_1, ..., t_j$ and the data should be independent, but does not provide any procedure or method for doing this. Breslow (1974) proposes to define the endpoints t_j of intervals as all observed failure times. More details and discussions concerning adequate choices for the time grid of the EP can be found in Gamerman (1991), Sahu *et al.* (1997) and Qiou *et al* (1999).

The main objectives of this article is to construct a bayesian Cox PH model,Cox (1972, 1975), under a piecewise constant function for baseline hazard $\lambda_0(t)$ with a gamma process as *prior* of cumulative hazard function, another objective is to review the methods and techniques used to determine the number and lengths of intervals in such piecewise function.

The important question we are trying to answer is the following: does the change in the number of intervals affect the parameters estimation and its standard deviations? what is the ideal number of intervals J? what would be the effect of the number of intervals on statistical tests based on the likelihood function like: Bayes Factor (see Raftery *et al*

(1995)), Likelihood ratio test(LR) (see Wilks (1938)), *and* even on the information criteria (BIC, DIC, AIC) (see Schwarz (1978)). We propose a method to determine the number of intervals by using the index of Huntsberger Cauvin *et al* (1987) and a *k-means* algorithm Gan*et al* (2007), to determine the lengths of theses intervals. By using two real data sets, we illustrate the estimation process via the **R** software.

This article is organized as follows: the Bayesian estimation process of Cox PH model is introduced in Section (2). The methodologies used to determine the number of intervals is discussed in Section (3). The Estimation of Cox PH model is illustrated with the analysis of a real data sets, and a discussion about the results are drawn in Section (4).

2. Bayesian Estimation of Cox Proportional hazard model

Suppose we have N subjects. Let $T_i = min(Y_i, C_i)$ be the observed time for the ith subject, with Y_i is the potential failure time and C_i is the potential censoring time. Let δ_i be the survival indicator, with $\delta_i = 1$ if $T_i = Y_i$ and $\delta_i = 0$ otherwise. Elementary, the data set can be denoted as pairs $(t_i, \delta_i)), i = 1, 2, ..., N$.

A standard formula of a proportional hazard model (PH) is given by :

$$\lambda(t \setminus x, \beta) = \lambda_0(t)\, \mathcal{J}(\beta, x_i) \tag{1}$$

where $\lambda_0(t)$ is the baseline hazard function when $\mathbf{x_i}$ takes value $\mathbf{x_i^0}$ such that $\mathcal{J}(\beta, x_i^0) = 1$, $\mathcal{J}(.)$ is a positive function. In the Cox model (see Cox (1972)), $\mathcal{J}(.)$ is an exponential function. Using the survival functions , the likelihood function of the model is proportional to :

$$\begin{aligned}
\mathcal{L}(X, \beta, \lambda_0) &= \prod_{i=1}^{n} f(t \setminus x_i, \beta)^{\delta_i}\, S(t \setminus x_i, \beta)^{(1-\delta_i)} \\
&= \prod_{i=1}^{n} [\lambda(t \setminus x_i, \beta)\, S(t \setminus x_i, \beta)]^{\delta_i} S(t \setminus x_i, \beta)^{(1-\delta_i)} \\
&= \prod_{i=1}^{n} \lambda(t \setminus x_i, \beta)^{\delta_i}\, S(t \setminus x_i, \beta) \\
&= \prod_{i=1}^{n} \left[\exp^{[\beta, X_i]}\ \lambda_0(t)\right]^{\delta_i} \left[S_0(t)\right]^{\exp^{[\beta, X_i]}}
\end{aligned} \tag{2}$$

By dividing and multiplying the equation (2), by : $\sum_{j \in R(t_i)} \exp^{[\beta, X_j]}$, we find :

$$\mathcal{L}(X, \beta, \lambda_0) = \prod_{i=1}^{n} \left[\frac{\left[\exp^{[\beta, X_i]}\ \lambda_0(t)\right]}{\sum_{j \in R(t_i)} \exp^{[\beta, X_j]}}\right]^{\delta_i} \left[S_0(t)\right]^{\exp^{[\beta, X_i]}} \times \left[\sum_{j \in R(t_i)} \exp^{[\beta, X_j]}\right]^{\delta_i}, \tag{3}$$

where $R(t_i)$ is a set of subjects at risk on t_i, which gives.

$$\mathcal{L}(X,\beta,\lambda_0) = \prod_{i=1}^{n}\Big[\frac{\exp^{[\beta,X_i]}}{\sum_{j\in R(t_i)}\exp^{[\beta,X_j]}}\Big]^{\delta_i} \tag{4}$$

We can see $\mathcal{L}(X,\beta,\lambda_0)$ as a conditional Likelihood. Under the condition of uncensored data and explanatory X_i variables that do not depend on time, Kalbfleisch (1978) have shown that the Eq.(4) is a marginal likelihood function of times $T_{(i)}$.

For more details about survival analysis and estimation methods, the references Collet (2014),Cox and Oakes (1984),Fleming and Harrington (2011) are of real interest.

2.1. Baselien Hazard Models

As explained in the section 1, Cox (1972, 1975) considers $\lambda_0(t)$ as a nuisance parameter when formulating the likelihood function. In Bayesian analysis, we must model this function. Several methods have been proposed for modeling the baseline hazard function $\lambda_0(t)$:
(a) - Parametric method: We could model the $\lambda_0(t)$ as a parametric hazard function, (e.g): Weibull, Gamma,.etc. For instance, for the Weibull, $\lambda_0(t)$ is defined as:

$$\lambda_0(t\ \alpha,\gamma) = \gamma\,\alpha\,t^{\alpha-1},$$

where the parameter α is a shape parameter and γ is a scale parameter.

(b) - Piecewise Exponential function : We assume that the hazard of occurrence of an event is *piecewise constant value exp*(λ_j) on the disjoint intervals of $0 < t_1 < t_2 < \ldots < t_J < t^*$, with $t^* = \max(t_j)$, $j = 0, ..., K$ and that we observe a survival time t_i in one of these intervals $(t_{j-1}, t_j]$.

$$\lambda_0(t) = \sum_{j=0}^{K}\lambda_j\,\mathbb{I}_{[t_j,t_{j+1}[}\,(t). \tag{5}$$

The survival function implied by this hazard and the corresponding likelihood is:

$$S(t) = \exp\big[-e^{x_i\beta}\sum_{j=1}^{k-1}e^{\lambda_j}(t_j - t_{j-1}) - (t-t^*)\big]$$

for $t \in (t_1, t^*]$, and the x_i's and β are the covariates and the regression parameters respectively.

2.2. Bayesian formulation of Cox PH model

One of the major problems of the Piecewise-Constant Exponential (PCE) models is to determine the number of time intervals (classes) to use. The number of time intervals must be determined by the analyst. First, let us begin by constructing the likelihood of the model

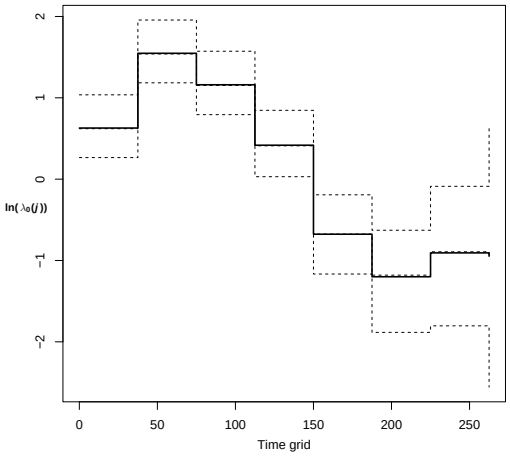

Fig. 1. An example of a piecewise function for baseline hazard

illustrated in Eq.(2), by using the piecewise constant method cited above, in Eq.(5). So, the ith term of the Eq.(2) is :

$$\begin{aligned}\mathcal{L}_i(X,\beta,\lambda_0(t)) &= \left[\exp^{[\beta,X_i]}\ \lambda_0(t)\right]^{\delta_i}\left[S_0(t)\right]^{\exp^{[\beta,X_i]}}\\ &= \left[\exp^{[\beta,X_i]}\ \lambda_0(t)\right]^{\delta_i}\left[\exp(-\Lambda_0(t_i))\right]^{\exp^{[\beta,X_i]}}\\ &= \left[\exp^{[\beta,X_i]}\ \lambda_0(t)\right]^{\delta_i}\ \exp\left[-\exp^{[\beta,X_i]}\ \Lambda_0(t_i)\right] \qquad (6)\end{aligned}$$

Now, we now integrate the equation 5 in this likelihood to get

$$\left[\exp^{[\beta,X_i]}\ \lambda_z\right]^{\delta_i}\ \exp\left[-\exp^{[\beta,X_i]}\ \sum_{j=1}^{z}\phi[i,j]\lambda_j\right], \qquad (7)$$

with,

$$\phi[i,j] = \begin{cases} t_{j+1}-t_j & \text{si } t_i \geq t_{j+1}\\ t_i - t_j & \text{si } t_i \in [t_j, t_{j+1}[\\ 0 & \text{si } t_i < t_j. \end{cases}$$

And z : is the maximum index of sub-survival interval (or tracking) of the individual i. Then, an indicator function is constructed, denoted $\psi(i,j)$, which identifies the sub-interval occurring at the event of interest from the individual general i,

$$\psi[i,j] = \begin{cases} 1 & \text{if } t_i \in [t_j, t_{j+1}[, \text{ and } \delta_i = 1.\\ 0 & \text{if else .} \end{cases}$$

For a particular individual $\underline{i}$, $\psi[\underline{i},z]=1$ indicates the occurrence of this event, $(i.e)$, $\psi[i,j]=0$, if $j \neq z$, and the individual $\underline{i}$ is *censored*. Returning to the Eq.(7), the form of the likelihood function of i becomes:

$$\mathcal{L}_i(X,\beta,\lambda) = \prod_{j=1}^{z} \left[\exp^{[\beta,X_i]}\ \lambda_j\right]^{\psi[i,j]} \exp\left[-\exp^{[\beta,X_i]}\ \phi[i,j]\lambda_j\right]. \tag{8}$$

As final step of reformulation, we write $\boldsymbol{\Upsilon}[\mathbf{i},\mathbf{j}] = \exp\left[-\exp^{[\beta,X_i]}\ \phi[i,j]\lambda_j\right]$, a condition of proportionality between this quantity and $\left[\exp^{[\beta,X_i]}\ \lambda_j\right]$. replacing $\boldsymbol{\Upsilon}[\mathbf{i},\mathbf{j}]$ in the likelihood, we obtain:

$$\mathcal{L}_i(X,\beta,\lambda) \propto \prod_{j=1}^{z} \left(\boldsymbol{\Upsilon}[\mathbf{i},\mathbf{j}]\right)^{\psi[\mathbf{i},\mathbf{j}]} \exp\left(-\boldsymbol{\Upsilon}[\mathbf{i},\mathbf{j}]\right). \tag{9}$$

We can clearly see that this likelihood is that of the variable $\psi[i,j]$ which follows a *Poisson distribution*, $\psi[i,j] \rightsquigarrow \mathcal{P}(\boldsymbol{\Upsilon}[\mathbf{i},\mathbf{j}])$, $j = 1,...,z$. Finally, the likelihood for all individuals is defined by:

$$\mathcal{L}(X,\beta,\lambda) \propto \prod_{i=1}^{n} \mathcal{L}_i(X,\beta,\lambda) = \prod_{i=1}^{n}\prod_{j=1}^{z} \left(\boldsymbol{\Upsilon}[\mathbf{i},\mathbf{j}]\right)^{\psi[\mathbf{i},\mathbf{j}]} \exp\left(-\boldsymbol{\Upsilon}[\mathbf{i},\mathbf{j}]\right). \tag{10}$$

For a more detailed introduction to the use of counting processes in survival models, the author is directed to Fleming and Harrington (2011) and to Ibrahim *et al.* (2001), the later being a main reference for Bayesian inference of survival models. An interesting study was conducted by Chen, *et al* (2014), in which another reformulation of the likelihood of the Cox model is proposed.

2.3. Priors for β_i and baseline hazard $\lambda_j(t)$

Priors for β_i : Generally, and in the context of the Bayesian estimation of survival models (see Ronald *et al.* (2011),Achcor*et al.* (1989)), we take a *reference* prior, for β_j $(i.e)$:

$$\beta_j \quad \underrightarrow{ind} \quad \mathcal{N}(\beta_{0j}, \sigma^2_{\beta_j})$$

The choice of this type could be considered as an initial *a priori* distribution, of which a sensitivity analysis is recommended (see Ibrahim *et al.* (2001)).

Prior for baseline hazard λ_j : A useful parametric probability distribution, *as a prior* distribution for the baseline hazard λ_j is the gamma distribution (see Arjas and Gasbarra (1994)):

$$\lambda_j \sim \mathcal{G}(\alpha z_j, z_j),$$

with a gamma probability density,

$$f(\lambda_j) = (z_j)^{\alpha z_j} \frac{1}{\Gamma(\alpha z_j)} \lambda_j^{\alpha z_j - 1} e^{-(z_j)\lambda_j}, \alpha z_j, z_j > 0, \lambda_j \geq 0 \tag{11}$$

Ronald *et al.* (2011) proposes to integrate the length of subintervals of tracking $[t_j, t_{j+1}[$ when constructing the prior distribution $(i.e)$,

$$\lambda_j \sim \mathcal{G}\big((t_{j+1}-t_j)\alpha z_j, (t_{j+1}-t_j) z_j\big).$$

From the likelihood function, in Eq(9), and the *prior* of the parameters β_i of the explanatory variables X, from Formula (2.3), and from the basic random function $\lambda_0(t)$, in Eq.(11), the form of the likelihood function *a posteriori* of the Cox model is defined by:

$$\pi(\beta, \lambda_j \setminus x_i, t_i) \ \propto \ \mathcal{L}(t_i, x_i, \beta, \lambda)\, \pi(\beta)\, \pi(\lambda_j).$$

which leads to

$$\begin{aligned} \pi(\beta, \lambda_j \setminus x_i, t_i) \ \propto & \\ & \propto \ \Big(\prod_{i=1}^{n}\prod_{j=1}^{z} \big(\boldsymbol{\Upsilon}[\mathbf{i},\mathbf{j}]\big)^{\psi[\mathbf{i},\mathbf{j}]} \exp\big(-\boldsymbol{\Upsilon}[\mathbf{i},\mathbf{j}]\big) \\ & \times \ \exp\Big(\frac{-1}{2}(\beta-\beta_0)\Sigma_0^{-1}(\beta-\beta_0)\Big) \\ & \times \ \lambda_j^{\alpha z_j - 1} e^{-(z_j)\lambda_j}\Big) \end{aligned} \tag{12}$$

The estimation of the parameters and the posterior laws and the underlying survival functions is carried out by MCMC methods using the Gibbs algorithm (see Geman and Geman (1984)). For more details of compilation and construction of such algorithms, we refer the reader to Robert (2006),Jayanta *et al* (2006) and Bolstad (2004) to name a few.

3. Partition of time axis (grid)

Discretizing a quantitative variable consists in mathematically transforming a vector of real numbers into a vector of integers named *class indices*. That is why doing this transformation is said in common language *to make a division into classes*. In statistics, discretising in achieved in a mathematical transformations leading to classes, in naming and justifying the grouping into classes.

Barry and Hartigan (2008) use product partition models, which assume that the probability of any partition is proportional to a product of prior cohesions, one for each block in the partition. That procedure is based on the fact that the blocks the parameters in different blocks have independent prior distributions. In a Gastric cancer study, Gamerman (1991) defines a grid using 77 distinct failure times, in which he suggested a grid, namely a $J = 30$ knot grid with $a = 0, 20, 40, 60, ..., 200, 250, 300, ..., 600, 700, ..., 1800$, where the grid is defined either by observed dead time or by a more aggregated partition, ideally such that each interval includes a balance of events among intervals. Gustafson*et al* (2003) propose a method of choosing nodes t_j located at $\big((j-1)/K\big)$ *quantiles* of the observed failure times.

Here, we propose a method based on Cauvin *et al* (1987)'s index, to determine the optimal number of intervals to achieve and for length intervals on Clustering analysis, by using the K-means algorithm ,Kaufman and Rousseeuw (1987),Gan*et al* (2007). The number of optimum classes to perform in a partition is always a function of the number of individuals observed (N). The Cauvin *et al* (1987)'s index allows to know the number of ideal classes for a distribution. It should be considered only as an indication. It is defined as below :

$$J = 1 + 3.32 \log_{10}(N),$$

with N = number of observations , J number of intervals.

In order to incorporate interval lengths, we used a *k-means* method which is a classical partitioning technique of clustering that clusters the data set of N objects into k classes (intervals, clusters) known *a priori* :

$$\mathcal{D}_m = \Big(\sum_{j=1}^{k} \sum_{i=1}^{N} w_i \big\| t_i^j - \bar{t}_j \big\|^p \Big)^{\frac{1}{p}}. \tag{13}$$

where the weights $0 \leq w_i \leq 1$ can be equal to 1 in the case of *unweighed distances*, or not in the case of *weighted distances*. This algorithm attempts to minimize the distance $\mathcal{D}_m$ between labeled points to be in a cluster k and a point designated as the center of that cluster $\bar{t}_j$. The parameter p is assumed is a positive integer. If $p = 1$, the distance is also known as the *Manhattan* norm. and if $p = 2$, the distance is also known as the *Euclidean* distance (Carugo (2010)). we use the later distance in our applications.

The steps of this algorithm are :

1. Clusters the data set (here survival time t_i into k groups where k is predefined.
2. Select k points at random as cluster centers.
3. Assign objects to their closest cluster center $\bar{t}_j$ according to the *Euclidean distance function.*
4. Calculate the centroid or mean of all objects in each cluster j.
5. Finally, Repeat steps **(2), (3)** and **4** until the same points t_i are assigned to each cluster j, in consecutive rounds.

4. Numerical Ilustrations and Discussion

Example 1:Gastric Cancer The data set used here is relative of a Gastric cancer , Stablein *et al* (1981), of survival times of patients with locally advanced, non respectable gastric carcinoma. The patients were either treated with chemotherapy plus radiation or chemotherapy alone.

Example 2: Lung Cancer : Survival in patients with advanced lung cancer from the North Central Cancer Treatment Group. Performance scores rate how well the patient can perform usual daily activities. Loprinzi*et al* (1994). The data set components are : inst : Institution code . time: Survival time in day status: censoring status 1=censored, 2=dead age: Age in years sex:Male=1 Female=2 ph.ecog:E COG performance score (0=good 5=dead) ph.karno: Karnofsky performance score (bad=0-good=100) rated by physician pat.karno: Karnofsky performance score as rated by patient meal.cal: Calories consumed at meals wt.loss: Weight loss in last six months.

It is shown that the log likelihood decreases when number of intervals J or censored sample size k increases for all the considered values. Unlikely, we have see a positive relationship with the Deviance information criterion (DIC). It is observable that no effect was found

Table 1. Estimation result of Cox model with changing of intervals number for Gastric cancer data set. J

J : Number of intervals	$\hat{\beta}$	$\hat{\sigma}_{\beta}$	Logliklihood	DIC
1	0.101	0.225	-144.8	287.8
5	0.111	0.225	-209.6	437.8
10	0.113	0.225	-261.3	563.1
15	0.113	0.225	-266.3	592.3
20	0.112	0.225	-293.5	665.1
25	0.113	0.225	-301.5	712.3
30	0.111	0.225	**-316.5**	772.5
50	0.112	0.225	-311.5	786.3
70	0.112	0.225	-312.5	836.5
90	0.112	0.225	-316.2	885.5

on the parameter $\hat{\beta}$ and its standard deviation $\hat{\sigma}_{\hat{\beta}}$. For the parameters of the explanatory variables, our interest is focused on the variation of the estimates and not on the statistical significance or the epidemiological interpretation (see 5). The point in blue corresponds to a number $J = 30$ of intervals. It is the same number used by Gamerman (1991) which coincides with the minimum of the log likelihood as in Table (1). By using our method, the estimated number of intervals, is $J \simeq 8$, which corresponds to a $LogLiklihood = -241.7$ and a $DIC = 509.3$. If we want to integrate the length of intervals in the prior distribution of baseline function $\lambda_j(t)$, the *K-means* method in Eq.(13) gives the following partition of time axis :

$[1; 315[, [315; 675[, [675; 1174[, [1174; 1366[, [1366; 1551[, [1551; 1690[, [1690; 1736[, [1736; 1736]$.
The point in green is an weighted average of Log-Likelihood, with the number of intervals as weights. it is correspond on a J.

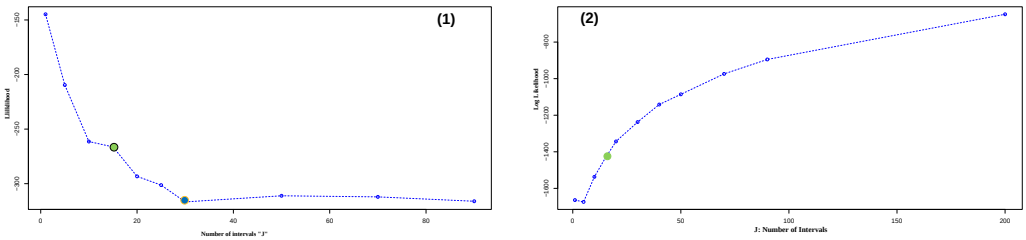

Fig. 2. Relation between the number of intervals J of piecewise function and marginal logliklihood function for **(1)** - Gastric cancer data. **(2)** - Lung Data

As a first remark, the number of intervals J has no effect on the results of the estimated parameters. Instead the variation of that number does logically have an impact on the likelihood function estimated models. This result, that is the variations of the log likelihood

function conditioned by the number of intervals, is an important trait of a choice of the number of intervals on the statistical tests based on the likelihood function, both in frequentist and the Bayesian approaches : see Raftery *et al* (1995)) for the Bayes Factor, Wilks (1938) for the likelihood ratio test(LR) and Schwarz (1978) for the criteria of statistical information. The described impact could more serious in the context of Model Selection and hypothesis testing.

Illustrative Example: For Bayesian, model selection and model criticism are extremely important inference problems. Sometimes these tend to become much more complicated than estimation problems. Without loss of generality, we may focus on two models from which we want to select the best one by using the DIC criterion. The idea is that models with smaller DIC should be preferred to models with larger DIC.

$$\begin{cases} H_0 : \text{data set } (t_0, t_1, ..., t_n) \sim \text{Bayesian Weibull Survival model }, \mathcal{W}(\alpha, \gamma). \\ H_1 : \text{data set } (t_0, t_1, ..., t_n) \sim \text{Cox PH model under PEM as baseline hazard function (5)} . \end{cases}$$

We have noticed that the volatility of log-likelihood $\mathcal{LL}$ values and the DIC criterion, depending on the number of intervals J chosen for the piecewise function. For example if the value DIC in the Weibull model is 2500, and that for the same modelling structure (linear predictor); therefore, (implicitly) and if we make an arbitrary choice of the number of intervals J, we can accept and reject the PEM model.
([1])

Table 2. Estimation result of Cox model with according to the number of intervals J , for Lung cancer data set.

N	intercept	age	sex	ph.ecog	ph.karno	pat.karno	$\mathcal{LL}$ (,)	DIC
1	-5.35	0.011	-1.01	1.31	0.02	-0.012	-1664	3229
5	-4.75	0.002	-0.956	0.914	0.005	-0.013	-1674	3255
10	-3.91	0.001	-0.988	0.914	0.005	-0.013	-1537	2988
15	-4.21	0.001	-0.991	0.914	0.005	-0.012	-1435	2791
20	-4.27	0.001	-0.992	0.991	0.005	-0.012	-1343	2619
30	-4.53	0.001	-0.996	0.907	0.005	-0.012	-1237	2425
40	-4.58	0.001	-0.991	0.912	0.005	-0.014	-1141	2255
50	-4.73	0.001	-0.994	0.912	0.005	-0.012	-1085	2159
70	-4.86	0.001	-0.991	0.912	0.005	-0.013	-974	1979
90	-4.99	0.001	-0.992	0.916	0.005	-0.013	-895	1836
200	-5.34	0.001	-0.993	0.915	0.005	-0.012	-648	1434

Noting that despite the variety of methods proposed for estimating the number intervals and their lengths of time grid, the used methods remain just an indicative tool in the in the Cox PH models or Piecewise Constant Exponential Models (PEM). Because the

[1] An advantage of DIC over other criteria in the case of Bayesian model selection is that the DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation.

nature of the considered phenomena in survival analysis are are heterogeneous, in the clinical field, in Biological studies, in Actuarial Sciences, in Reliability, etc. Some phenomena are characterized by a high risk rate $\lambda_0(t)$ at the beginning of follow-up which became stable afterwards. The opposite happens for others. This behaviour is not easy to control and to mathematically analyse. Therefore, it is highly recommend to appeal to a statistician, a specialist of area, to determine the limit number of intervals and even the underlying lengths is necessary. We have seen that an arbitrary choice of this number could lead to a poor modeling. In that context, the case of Gamerman (1991) was well identified in our study, see the line 8 in Table 1.

So for this reason, when trying to get a good fit for the hazard function $\lambda(t \setminus x_i, \delta_i)$, we should fix the number of split points to some reasonable number and let the locations and heights on each interval to vary. This number has to be a compromise between an adequately modelling the shape of the hazard and ensuring that there is enough data in each interval to get a proper fit. This issue has been one of the biggest challenges of working, particularly with the EMP.

References

Kalbfleisch JD. Non-parametric bayesian analysis of survival time data.1978. *Journal of the Royal Statistical Society B.*

Seong W. Kim. and Joseph G. Ibrahim. 2000.On Bayesian Inference for Proportional Hazards Models Using Noninformative Priors. *Lifetime Data Analysis*, 6, 331–341.

Sahu, S. K., Dey, D. K., Aslanidu, H. and Sinha, D.1997.A weibull regression model with gamma frailties for multivariate survival data. *Lifetime Data Analysis*, 3: 123–137.

Gamerman, D. (1991) Dynamic Bayesian models for survival data. *Applied Statistics*, 40, 63–79.

Ronald .C, Wesley .J. Bayesian Ideas and Data Analysis : An Introduction for Scientists and Statisticians. 2011.Taylor Francis Group, New york.

Arjas, E. and Gasbarra, D.1994. Nonparametric Bayesian inference for right-censored survival data, using the Gibbs sampler. *Statistica Sinica*, 4, 505–524.

Demarqui, F. N., Loschi, R. H. and Colosimo, E. A. 2008. Estimating the grid of time-points for the piecewise exponential model, *Lifetime Data Anal* 14: 333–356.

Achcor, J.A. and Bolfarine, H. 1989 Constant hazard against a change-point alternative:A Bayesian approach with censored data, *Comm. Stat. - Th. & Meth*, 18, 3801-3819.

Basu, A.P., Ghosh, J.K. and Joshi, S.N. (1988). On estimating change point in a failure rate. S*tatistical decision theory and related topics IV*, 2, Springer-Verlag, New York, 239-252.

Daniel Barry; J. A. Hartigan,A Bayesian Analysis for Change Point Problems.(1993).*Journal of the American Statistical Association*, Vol. 88, No. 421, pp. 309-319.

D. R Cox. Regression models and life-tables. (1972).*J. Roy. Statist.Soc*, (34) :187–220, 1972.

D. R. Cox. Partial likelihood.1975. *Oxford University Press on behalf of Biometrika Trust.*

Collett, D.: Modelling Survival Data in Medical Research, 3rd edn. 2014. Chapman and Hall/CRC, Boca Raton.

Cox, D.R., Oakes, D.: Analysis of Survival Data. 1984.Chapman and Hall/CRC, London.

Gan G., Ma C., Wu J.,2007. Data Clustering – Theory, Algorithms and Applications .2007.SIAM.

Gustafson.P.Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustment.2003.Chapman . Hall/CRC.London.

Fleming, T.R., Harrington, D.P.: Counting Processes and Survival Analysis.2011. Wiley, Hoboken.

Ibrahim,J.G,M.H.Chen, and D.Sinha.2001. Bayesian Survival Analysis.2001. Springer-Verlag,New York.

.A Bayesian Approach for the Cox Proportional Hazards Model with Covariates Subject to Detection Limit.2014. *Int J Stat Med Res*.3(1) . :32-43.

Stablein DM, Carter J, Novak JW. 1981. Analysis of Survival Data with Nonproportional Hazard Functions. *Controlled Clinical Trials* 2:149-159.

Christian P. Robert.2006. Le choix bayésien Principes et pratique , Springer-Verlag, Paris, France.

Jayanta K. Ghosh ,Mohan Delampady .(2006). *An Introduction to Bayesian Analysis Theory and Methods*,Springer.

William .M.Bolstad. Introduction to Bayesian Statistics.2004. Wiley, New Zealand.

Geman.S. and Geman.D. Stochastic relaxation, gibbs distributions,and the bayesian restoration of images.1984. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6) :721–741.

Loprinzi CL. Laurie JA. Wieand HS. Krook JE. Novotny PJ. Kugler JW. Bartel J. Law M. Bateman M. Klatt NE. *et al*.1994. Prospective evaluation of prognostic variables from patient-completed questionnaires. North Central Cancer Treatment Group. *Journal of Clinical Oncology*. 12(3):601-7 .

Cauvin.C, Reymond.H, Serradj.A.1987. Discrétisation et représentation cartographique, Montpellier, GIP RECLUS.

Friedman, M. 1982. Piecewise Exponential Models for Survival Data with Covariates. *The Annals of Statistics*, 10(1), pp. 101-113.

Qiou, Z., Ravishanker, N. and Dey, D. 1999. Multivariate survival analysis with positive stable frailties, *Biometrics* 55, 637–644.

Wilks, S. S. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses.1938. *The Annals of Mathematical Statistics*. 9: 60–62. doi:10.1214/aoms/1177732360.

Kass, Robert E. Raftery, Adrian E. 1995, Bayes Factors, *Journal of the American Statistical Association*, 90 (430): 773–795, doi:10.2307/2291091.

Schwarz, Gideon E. 1978.Estimating the dimension of a model. *Annals of Statistics*, 6 (2): 461–464, doi:10.1214/aos/1176344136.

Kalbfleisch .John. D. , Ross L. Prentice .2011. The Statistical Analysis of Failure Time Data. Print ISBN:9780471363576.DOI:10.1002/9781118032985.John Wiley / Sons, Inc.

Kaufman, L. Rousseeuw, P.J. 1987. Clustering by means of Medoids, in Statistical Data Analysis Based on the L_1 and Related Methods, edited by Y. Dodge, North-Holland, 405–416.

O. Carugo, F. Eisenhaber (eds.),2010. Data Mining Techniques for the Life Sciences, Methods in Molecular Biology. , Humana Press, a part of Springer Science Business Media, LLC. DOI10.1007/978-1-60327-241-4_10

Breslow, N. E. 1974. Covariance analysis of censored survival data, Biometrics 30: 89–99.

5. Appendix

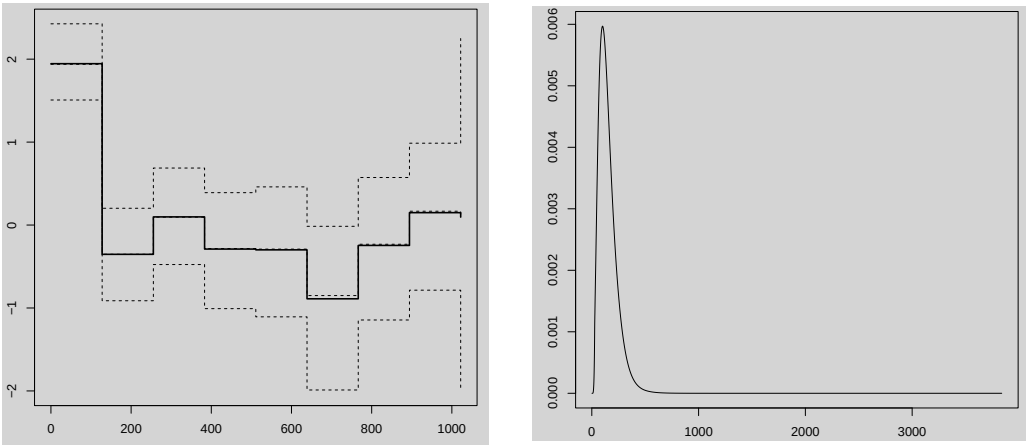

Fig. 3. Posterior baseline hazard function and its posterior precision for Gastric cancer data.

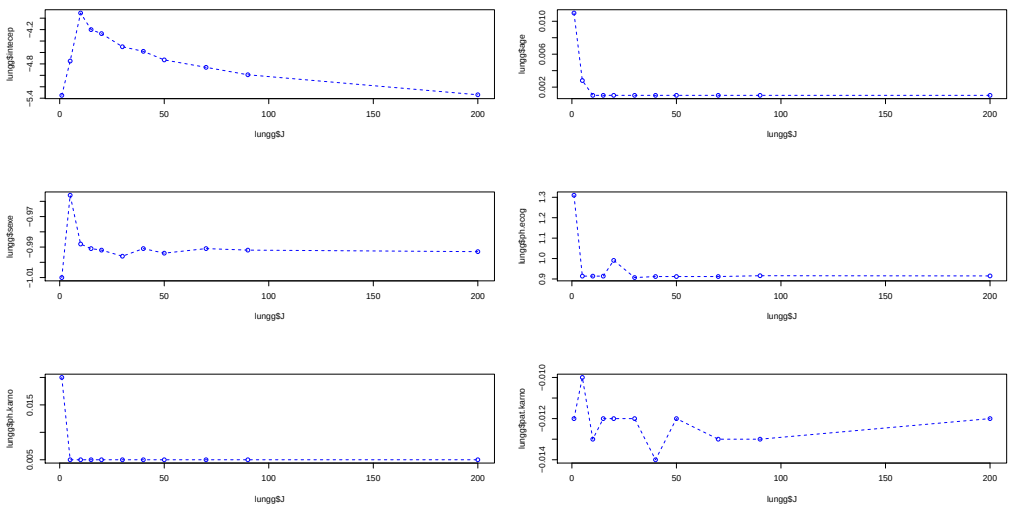

Fig. 4. Variations of estimates parameters according to the Number of intervals, in Lung cancer data set.