Proportional hazard models with an unknown link function : How to estimate the impact of time and covariate when using Cox model ?

Baptiste Aussel and Albin Cintas

December 2021

Contents

1 Introduction

Analysis of survival time can be very useful in several fields such as manufacturing or medicine. These kind of analysis are considered when it comes to evaluate the quality of a manufactured product or to estimate the remaining lifetime of someone that is sick. For sure time is a main criteria in these studies, but it is not the only one, environmental variables can also be part of the influential factors.
In this project we will study the approach of Jianqing Fan, Irène Gijbels and Martin King concerning a probabilistic estimation of survival functions written in the article "Local Likelihood and Local Partial Likelihood in Hazard Regression".
Indeed, Cox has proposed a model which distinguishes the effects of time and covariate on reliability function that we aim to estimate in this project. To do that we will firstly generate data with Cox modelling and different values of covariate. Then we will implement a model that aims at estimating the parameters of our Cox modelling. Finally we will apply the better method to a dataset to modelize its reliability function.

2 Cox model

In this report, we define T as the random variable which denotes times of death. Firstly it is important to define some functions used in reliability studies :

$$R(t) = P(T > t) \text{ : } \textit{reliability function}$$
$$\lambda(t) = \lim_{h \to 0} \frac{1}{h} P(T \in [t, t+h] | T > t) \text{ : } \textit{hazard rate}$$
$$\Lambda(t) = \int_0^t \lambda(s) ds \text{ : } \textit{cumulative hazard function}$$
$$R(t) = exp(-\Lambda(t)) \text{ : } \textit{relation between R and } \Lambda$$

Also known as the proportional hazard model, cox model define the hazard rate as a function which separates time and covariate impacts. Let us define X as the random variable which evaluates the covariate. We have the following relation according to Cox :

$$\lambda(t|x) = \lambda_0(t, \theta) \times exp(\psi(x))$$

with $\psi(x)$ a function of the covariate and $\lambda_0(t, \theta)$ a distribution which depends on the parameter θ. This distribution is called baseline as it is equal to $\lambda(t, 0)$.
A property of this model is that the hazard ratio between two individuals does not depend on the time :

$$\frac{\lambda(t|x_1)}{\lambda(t|x_2)} = exp(\psi(x_1) - \psi(x_2))$$

objective : In this project we aim at estimating $\hat{\lambda_0}(t, \theta)$ and $\hat{\psi}(x)$.

3 Data generation

We decide to study mainly the Weibull distribution concerning the λ_0 function:

$$\lambda_0(t) = \frac{\beta}{\alpha}(\frac{t}{\beta})^{\beta-1}$$

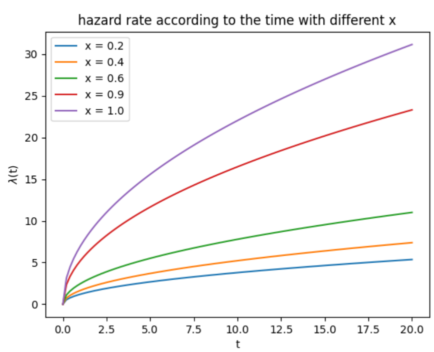
We design a ψ function which verifies $\psi(0) = 0$:

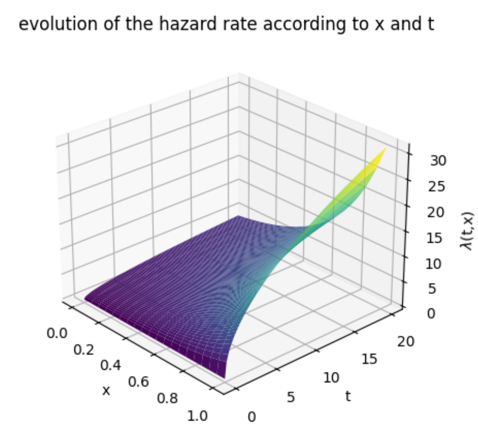
$$\psi(x) = x^2 + x$$

We have then a hazard rate function defined by :

$$\lambda(t|x) = \frac{\beta}{\alpha}(\frac{t}{\beta})^{\beta-1} exp(x^2 + x) = \lambda_0(t, theta) \times exp(\psi(x))$$

This hazard rate is increasing according to the x :





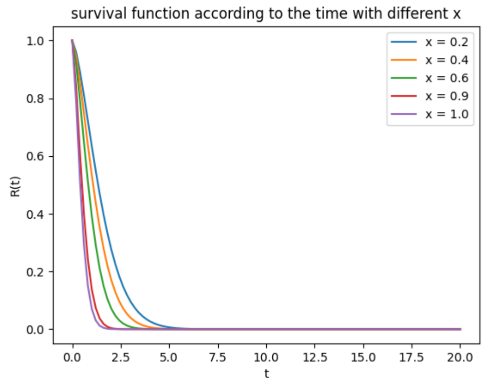
We can deduce the cumulative hazard rate knowing a covariate :

$$\Lambda(t|x) = \Lambda_0(t, theta) \times exp(\psi(x))$$

And then the reliability (survival) function :

$$R(t|x) = exp(-\Lambda_0(t, theta) \times exp(\psi(x)))$$

As defined we have the following reliability function in theory :



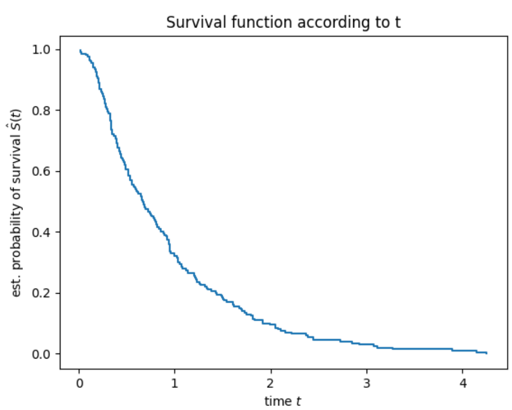
The process used to generate data consists in first creating a set of values randomly chosen for the reliability function, and then inverse the function to obtain a set of death time. To do that, we generate a random uniform set :

$$R \sim U_{[0,1]}$$

Then we have :

$$R_i = exp(-\Lambda_0(t_i, theta) \times exp(\psi(x)))$$
$$\Leftrightarrow$$
$$t_i = \alpha \times [-ln(R_i) \times exp(-\psi(x))]^{\frac{1}{\beta}}$$

We obtain generated a 200 individuals data set, which gives a reliability functions as follow :



4 Estimation with local likelihood described by the article

4.1 Presentation

In the article from Fan Jianqing, Irene Gijbels and Martin King, we are interested in the estimations of λ_0 and ψ.
This estimation approach come from a likelihood study that we define as :

$$logL = \sum_{i=1}^{n}[\delta_i(log\lambda_0(T_i;\theta) + \psi(X_i)) - \Lambda_0(T_i,\theta)exp(\psi(X_i)]$$

This likelihood leads us to the concept of the local likelihood which consists in watching the likelihood on a neighborhood of a covariate, and then approximate by a polynomial function.
We have then

$$\psi(X) \sim \boldsymbol{X}^T\gamma$$

with $\gamma = \{\psi(x), \psi'(x), \psi^{(2)}(x)/2, ...\}$ until the p degree.
And $\boldsymbol{X} = \{1, X_i - x, (X_i - x)^2, ...\}$ until the p degree.
If we introduce a kernel which will add weights according to the distance from the neighbor, we can deduce the following local log likelihood :

$$l_n(\gamma,\theta) = \frac{1}{n}\sum[\delta_i(log\lambda_0(T_i,\theta) + \boldsymbol{X_i}^T\gamma) - \Lambda_0(X_i,\theta)exp(\boldsymbol{X_i}^T\gamma)]K_h(X_i - x)$$

If we derive these terms according to γ and θ and we solve these derived equals to 0, we can estimate our γ and θ. We can also play on different parameters such as p, h of the kernel and K_h itself and see how the estimation evaluates, and that is what we are going to see afterwards.

4.2 first approach : p = 2, K_h Gaussian and h = 2

In this section we are fixing the degree of p to 2.
We want to estimate $\gamma = \{\psi(x), \psi'(x), \psi^{(2)}(x)/2\}$ $\qquad \psi(x) = x^2 + x$
Which equals here to

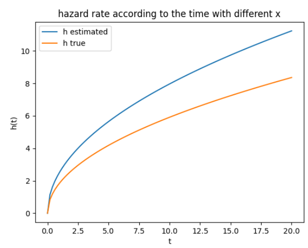
$$\gamma = \{x^2 + x, 2x + 1, 1\}$$

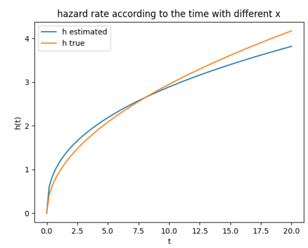
And

$$\boldsymbol{X} = \{1, X_i - x, (X_i - x)^2\}$$

where $\boldsymbol{X_i} \sim \mathcal{U}([0;1]) \qquad \forall i \in 1 : n$

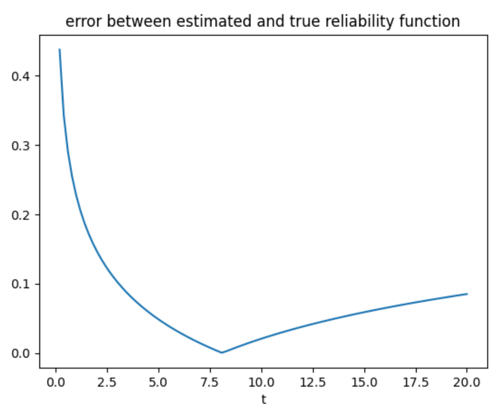
We can show the results, as there is hazard in data generation, we remark that it impacts the results :





We can see here that the estimations can be very good if the data fits well the model. From now on, we will keep the data set from the right estimation and evaluate the error as :

$$error = \frac{|\hat{\lambda} - \lambda_{true}|}{\lambda_{true}}$$



4.3 variation of p, K_h and h

We will now do a grid search and modify the three parameters to see what is the best combination with our data set. To evaluate our results, we will compute the norm of the error we have just plot above ($||\frac{|\hat{\lambda}-\lambda_{true}|}{\lambda_{true}}||$).

We organised the research as follow :

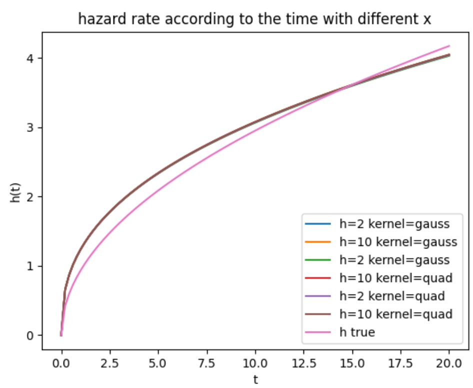
- We will test 3 kernels : gaussian, quadratic and Epanechnikov.

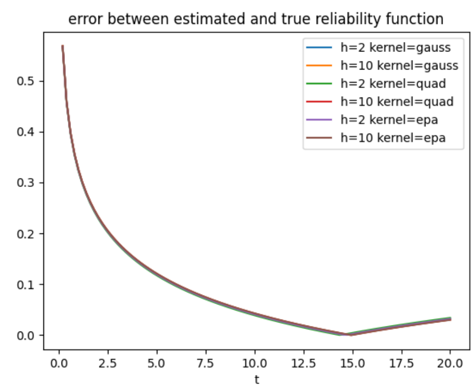
  $K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}$

$K(u) = \frac{15}{16}(1-u^2)^2 \mathbb{1}_{\{|u|\leq 1\}}$
$K(u) = \frac{3}{4}(1-u^2) \mathbb{1}_{\{|u|\leq 1\}}$

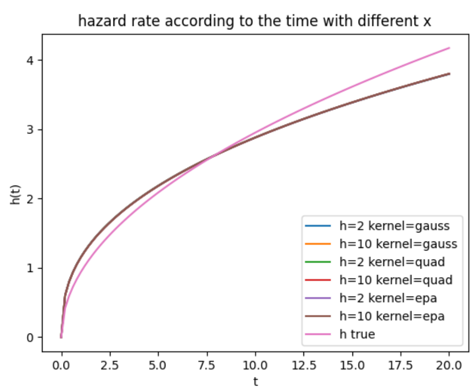
- We will test all of them for h = 2 and h = 10
- We will test all of these combinations for p in 1, 2, 3

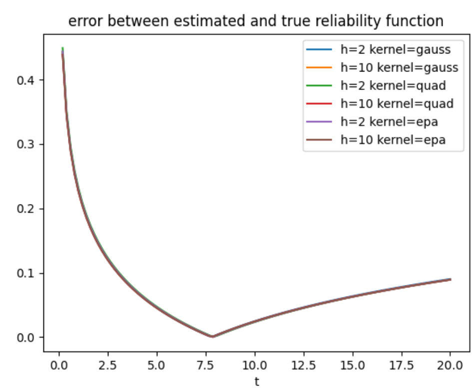
We have plotted all the hazard rate found and its error from the true value classified by p:
For p = 1 :



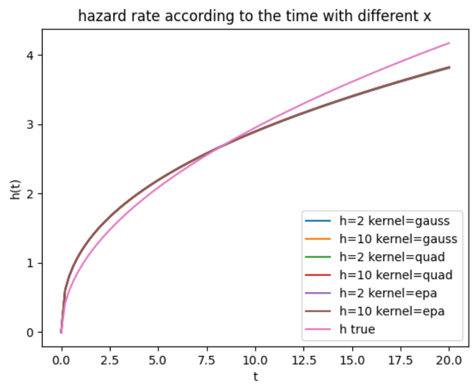


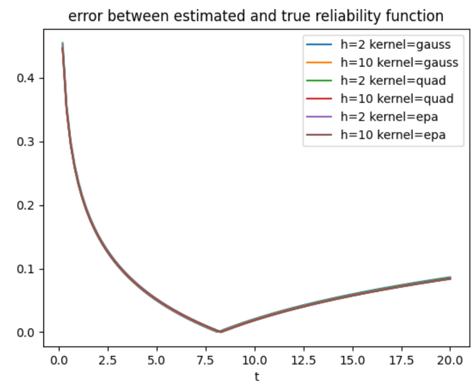
For p = 2 :





For p = 3 :





In these displays we can see that when we fix the p parameter, there is not so much changes between different kernels and different h.
Now to evaluate and compare the error between all of these instances, we have built up the following table which relates all of the errors for our use cases :

Kernel/parameters	p=1	p=2	p=3
Gaussian h=2	1.356	1.002	1.007
Gaussian h= 10	1.362	1.001	1.007
Epanechnikov h=2	1.347	1.013	1.021
Epanechnikov h=10	1.362	0.994	1.003
Quadratic h=2	1.354	1.003	1.012
Quadratic h=10	1.362	0.993	1.002

We keep for the following parts the quadratic kernel, with h = 10 and p = 2. We saw indeed that from p = 1 to p = 2, there are significant improvements, but these improvements doesn't translate from p = 2 to p = 3, where there is no more change.

4.4 add of censure

We decided to add test different % of censure on our data set. We performed experiments and deduced that the more we add censure, the less the estimator performs. This is a logical issue.

Kernel parameters/ uncensored rate	90%	80%	70%	60%	50%
Quadratic h=10	0.72	1.604	2.553	4.484	5.150

5 Application on a data set

For the application, we used a data set called NCCTG lung Cancer. This data set reports the death time of a sample of people who were sick according to different parameters (their age, sex, health...).

We performed the method of local likelihood on the data set, but it did not seem to be efficient. Indeed, the algorithm don't reach the optimal point. An hypothesis is that the assumption we took of considering λ_0 as a Weibull law is not veryfied here, and then we cannot obtain any result.

6 Conclusion

As conclusion we can say that the Cox model is efficient, if λ_0 follows a weibull law to evaluate the relation between the survival function and a covariate.

The more we put censure on our data the less our model will be efficient. With the different parameter we decided to study we found that the better model is given for the degree of the approximation of ϕ p=2 , h=10 and with the Epanechnikov kernel.

We also found that if the data set term related to time doesn't follow a Weibull law the results can be really bad.

However, the study field is very large and have a lot of different parameters, among which we can find a range of kernels. We tested some of them but we could not test them all, so maybe some other kernels with different parameters could give better results.

It would also be interesting to test other laws for the baseline to see if it has a big impact on real data sets.