Introduction to BGPhazard

José A. García-Bueno Emilio A. Morones Ishikawa and Luis E. Nieto-Barajas

2023-09-02

Contents

Abstract

We present `BGPhazard`, an R package which computes hazard rates from a Bayesian nonparametric view. This is achieved by computing the posterior distribution of a gamma or a beta process through a Gibbs sampler. The purpose of this document is to guide the user on how to use the package rather than conducting a thorough analysis of the theoretical results. Nevertheless, section 2 briefly discuss the main results of the models proposed by Nieto-Barajas and Walker (2002) and by Nieto-Barajas (2003). These results will be helpful to understand the usage of the functions contained in the package. In section 3 we show some examples to illustrate the models.

Introduction

Survival analysis focuses on studying data related to the occurrence time of an event. A typical function is the *survival function*, which in nonparametric statistics is estimated through the product limit estimator (Kaplan & Meier, 1958). This estimator is used as an approximation to the survival function. In some cases, its stair-step nature can return misleading estimators in a neighborhood of the steps: just before and after each we will have significant differences between estimators.

Lack of smoothness on nonparametric estimations give rise to methods whose outputs are smooth functions. One of many approaches is given by Nieto-Barajas and Walker (2002) for the survival function. The theory

behind these models combines Bayesian Statistics and Survival Analysis to obtain hazard rate estimates. Bayesian Statistics let us introduce previous knowledge to a data set to improve estimations. Nieto-Barajas & Walker (2002) estimate the hazard function in segments by introducing dependence between each other so that information is shared and a smooth hazard rate is obtained. We review the three models contained in the package: beta model for discrete data, gamma model for continuous data and cox-gamma model for continuous data in a proportional hazards model.

A consequence of using a nonparametric Bayesian model with a dependence structure is that the resulting estimators are smoother than those obtained with a frequentist nonparametric model.

Hazard rate estimation

In this brief review, we examine a generalization of the independent gamma process of Walker and Mallick (1997) –gamma model–; then, a generalization of the beta process introduced by Hjort (1990) –beta model– which is often used to model discrete failure times, and lastly, the proportional risk model extension to the gamma process that copes with explanatory variables that remain constant during time (Nieto-Barajas, 2003).

We provide nonparametric prior distributions for the hazard rate based on the dependence processes previously defined and we obtain the posterior distributions through a Bayesian update.

Markov beta and gamma prior processes

Let λ_k represent the gamma process and let π_k represent the beta process. Let θ_k represent both λ_k and π_k. For interpretation of the Markov model, the main priority is to ensure

$$E[\theta_{k+1}|\theta_k] = a + b\theta_k$$

A latent process $\{u_k\}$ is introduced in order to obtain $\{\theta_k\}$ from

$$\theta_1 \to u_1 \to \theta_2 \to u_2 \to \cdots$$

Gamma Process

Walker & Mallick (1997) consider $\{\lambda_k\}$ as independent gamma variables, $i.e.$, $\lambda_k \sim Ga(\alpha_k, \beta_k)$ independent for $k = 1, 2, ...$ Nieto-Barajas & Walker (2002) consider a dependent process for $\{\lambda_k\}$. They start with $\lambda_1 \sim Ga(\alpha_1, \beta_1)$ and take $u_k|\lambda_k \sim Po(c_k\lambda_k)$ and $\lambda_{k+1}|u_k \sim Ga(\alpha_{k+1} + u_k, \beta_{k+1} + c_k)$ and so on. These updates arise from the joint density

$$f(u, \lambda) = Ga(\lambda|\alpha, \beta)Po(u|c\lambda)$$

and so constitute a Gibbs type update. The difference is that they are changing the parameters (α, β, c) at each update so the chain is not stationary and marginally the $\{\lambda_k\}$ are not gamma.

However,

$$\mathrm{E}[\lambda_{k+1}|\lambda_k] = \frac{\alpha_{k+1} + c_k\lambda_k}{\beta_{k+1} + c_k}$$

If $c_k = 0$, then $P(u_k = 0) = 1$ and hence the $\{\lambda_k\}$ are independent gamma and we have the prior process of Walker and Mallick (1997).

An important result is that if we take $\alpha_k = \alpha_1$ and $\beta_k = \beta_1$ to be constant for all k, then the process $\{u_k\}$ is a Poisson-gamma process with implied marginals $\lambda_k \sim Ga(\alpha_1, \beta_1)$. One can note that if u_1 is Poisson distributed and $\lambda_2|u_1$ is conditionally gamma, then λ_2 is never gamma.

Beta Process

Nieto-Barajas and Walker (2002) start with $\pi_1 \sim Be(\alpha_1, \beta_1)$ and take $u_k|\pi_k \sim Bi(c_k, \pi_k)$, $\pi_{k+1}|u_k \sim Be(\alpha_{k+1} + u_k, \beta_{k+1} + c_k - u_k)$ and so on. These arise from a binomial-beta conjugate set-up, from the joint density

$$f(u, \pi) = Be(\pi|\alpha, \beta)Bi(u|c, \pi)$$

Clearly

$$\mathrm{E}[\pi_{k+1}|\pi_k] = \frac{\alpha_{k+1} + c_k\pi_k}{\alpha_{k+1} + \beta_{k+1} + c_k}$$

As with the gamma process, if we choose $c_k = 0$, then $P(u_k = 0) = 1$ and so the $\{u_k\}$ become independent beta and we obtain the prior of Hjort (1990).

A significant result is that if we take $\alpha_k = \alpha_1$ and $\beta_k = \beta_1$ to be constant for all k, then the process $\{u_k\}$ is a binomial-beta process with marginals $u_k \sim BiBe(\alpha_1, \beta_1, c_k)$. Moreover, the process $\{\pi_k\}$ becomes strictly stationary and marginally $\pi_k \sim Be(\alpha_1, \beta_1)$.

Prior to posterior analysis

We use the gamma and beta processes to define nonparametric prior distributions. In order to obtain $f(\theta|data)$, we introduce the latent variable u and so constitute a Gibbs update for $f(\theta|u, data)$ and $f(u|\theta, data)$. As we generate a sample from $f(\theta, u|data)$, we automatically obtain a sample from $f(\theta|data)$. Therefore, given u from $f(u|\theta, data)$, we can obtain a sample from $f(\theta, u|data)$ by simulating from $\theta \sim f(\theta|u, data)$.

It can be shown that

$$f(\theta|u, data) \propto f(data|\theta) \times f(\theta|u)$$

and

$$f(u|\theta, data) \propto f(data|\theta, u) \times f(u|\theta) \propto f(u|\theta)$$

because u and $data$ are conditionally independent given θ.

Gamma Process

Let T be a continuous random variable with cumulative distribution function $F(t) = P(T \leq t)$ on $[0, \infty)$. Consider the time axis partition $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$, and let λ_k be the hazard rate in the interval $(\tau_{k-1}, \tau_k]$, then the hazard function is given by

$$h(t) = \sum_{k=1}^{\infty} \lambda_k I_{(\tau_{k-1}, \tau_k]}(t)$$

So, the cumulative distribution and density functions, given $\{\lambda_k\}$, are $F(t|\{\lambda_k\}) = 1 - e^{-H(t)}$, $f(t|\{\lambda_k\}) = h(t)e^{-H(t)}$, where $H(t) = \int_0^t h(s)ds$. We also have that

$$f(\lambda_k|u_{k-1}, u_k) = Ga(\alpha_k + u_{k-1} + u_k, \beta_k + c_{k-1} + c_k)$$

Therefore, given a sample $T_1, T_2, ..., T_n$ from $f(t\{\lambda_k\})$, it is straightforward to derive

$$f(\lambda_k|u_{k-1}, u_k, data) = Ga(\alpha_k + u_{k-1} + u_k + n_k, \beta_k + c_{k-1} + c_k + m_k),$$

where n_k = number of uncensored observations in $(\tau_{k-1}, \tau_k]$, $m_k = \sum_i r_{ki}$, and

$$r_{ki} = \begin{cases} \tau_k - \tau_{k-1} & t_i > \tau_k \\ t_i - \tau_{k-1} & t \in (\tau_{k-1}, \tau_k] \\ 0 & \text{otherwise} \end{cases}$$

Additionally,

$$P(u_k = u | \lambda_k, \lambda_{k+1}, data) \propto f(u|\lambda) \propto \frac{[c_k(c_k + \beta_{k+1})\lambda_k\lambda_{k+1}]^u}{\Gamma(u+1)\Gamma(\alpha_{k+1}+u)}$$

with $u = 0, 1, 2, ...$ Hence, with these full conditional distributions, a Gibbs sampler is straightforward to implement in order to obtain posterior summaries.

We can learn about the $\{c_k\}$ by assigning an independent exponential distribution with mean ϵ for each c_k, $k = 1, .., K - 1$. The Gibbs sampler can be extended to include the full conditional densities for each c_k. It is not difficult to derive that a c_k from $f(c_k|u, \lambda, data)$ can be taken from the density

$$f(c_k|u_k, \lambda_k, \lambda_{k+1}) \propto (\beta_{k+1} + c_k)^{\alpha_{k+1}+u_k} c_k^{u_k} \exp\left\{-c_k\left(\lambda_{k+1} + \lambda_k + \frac{1}{\epsilon}\right)\right\}$$

for $c_k > 0$.

Dependence between c_k's can be introduced through a hierarchical model via assigning a distribution to $\epsilon \sim Ga(a_0, b_0)$. The update would be given by:

$$f(\epsilon|\{c_k\}) = Ga\left(\epsilon|a_0 + K, b_0 + \sum_{k=1}^{K} c_k\right)$$

where K is the number of intervals generated by the time axis partition. This hierarchical specification of the initial distribution for c_k let us assign a better value for c_k.

Simulating from this distribution is not so straightforward. However, we construct a hybrid algorithm using a Metropolis-Hastings scheme taking advantage of the Markov chain generated by the Gibbs sampling.

Beta process

Let T be a discrete random variable taking values in the set $\{\tau_1, \tau_2, ...\}$ with probability density function $f(\tau_k) = P(T = \tau_k)$. Let π_k be the hazard rate at τ_k, then the cumulative distribution and the density functions, given $\{\pi_k\}$, are $F(\tau_j|\{\pi_k\}) = 1 - \prod_{k=1}^{j}(1 - \pi_k)$ and $f(\tau_j|\{\pi_k\}) = \pi_j \prod_{k=1}^{j-1}(1 - \pi_k)$

The conditional distribution of π_k is

$$f(\pi_k|u_{k-1}, u_k) = Be(\pi_k|\alpha_k + u_{k-1} + u_k, \beta_k + c_{k-1} - u_{k-1} + c_k - u_k),$$

Thus, given a sample $T_1, T_2, ..., T_n$ form $f(\cdot|\{\pi_k\})$ it is straightforward to derive

$$f(\pi_k|u_{k-1}, u_k, data) = Be(\pi_k|\alpha_k + u_{k-1} + u_k + n_k, \beta_k + c_{k-1} - u_{k-1} + c_k - u_k + m_k),$$

where n_k = number of failures at τ_k, $m_k = \sum_i r_{ki}$ and

$$r_{ki} = \begin{cases} 1 & t_i > \tau_k \\ 0 & \text{otherwise} \end{cases}$$

Additionally,

$$P(u_k = u | \pi_k, \pi_{k+1}, data) \propto \frac{\theta_k^u}{\Gamma(u+1)\Gamma(c_k - u + 1)\Gamma(\alpha_{k+1} + u)\Gamma(\beta_{k+1} + c_k - u)}$$

with $u = 0, 1, ..., c_k$ and

$$\theta_k = \frac{\pi_k \pi_{k+1}}{(1 - \pi_k)(1 - \pi_{k+1})}$$

As before, obtaining posterior summaries via Gibbs sampler is simple.

We can learn about the $\{c_k\}$ via assigning each c_k an independent Poisson distribution with mean ϵ. The Gibbs sampler can be extended to include the full conditional densities for each c_k. A c_k from $f(c_k | u, \pi, data)$ can be taken from the density

$$f(c_k | u_k, \pi_k, \pi_{k+1}) \propto \frac{\Gamma(\alpha_{k+1} + \beta_{k+1} + c_k)}{\Gamma(\beta_{k+1} + c_k - u_k)\Gamma(c_k - u_k + 1)} [\epsilon(1 - \pi_{k+1})(1 - \pi_k)]^{c_k}$$

with $c_k \in \{u_k, u_k + 1, u_k + 2, ...\}$.

Dependence between c_k's can be introduced through a hierarchical model via assigning a distribution to $\epsilon \sim Ga(a_0, b_0)$. So the update would be given by:

$$f(\epsilon | \{c_k\}) = Ga\left(\epsilon | a_0 + K, b_0 + \sum_{k=1}^{K} c_k\right)$$

where K is the number of discrete values in random variable T.

Cox-gamma model

Differing from most of the previous Bayesian analysis of the proportional hazards model, Nieto-Barajas (2003) models the baseline hazard rate function with a stochastic process. Let T_i be a non negative random variable which represents the failure time of i and $Z_i = (Z_{i1}, ..., Z_{ip})$ the vector containing its p explanatory variables. Therefore, the hazard function for individual i is:

$$\lambda_i(t) = \lambda_0(t) exp\{Z_i'\theta\}$$

where $\lambda_0(t)$ is the baseline hazard rate and θ is regression's coefficient vector. The cumulative hazard function for individual i becomes

$$H_i(t) = \sum_{k=1}^{\infty} \lambda_k W_{i,k}(t, \lambda),$$

where,

$$W_{i,k}(t, \theta) = \begin{cases} (\tau_k - \tau_{k-1}) \exp\{Z_i'\theta\} & t_i > \tau_k \\ (t_i - \tau_{k-1}) \exp\{Z_i'\theta\} & t_i \in (\tau_{k-1}, \tau_k] \\ 0 & otherwise \end{cases}$$

Given a sample of possible right-censored observations where $T_1, ..., T_{n_u}$ are uncensored and $T_{n_u+1}, ..., T_n$ are right-censored, the conditional posterior distributions for the parameters of the semi-parametric model are:

- $f(\lambda_k | u_{k-1}, u_k, data, \theta) = Ga(\lambda_k | \alpha_k + u_{k-1} + u_k + n_k, \beta_k + c_{k-1} + c_k + m_k(\theta))$

- $P(u_k = u | \lambda_k, \lambda_{k+1}, data) \propto f(u|\lambda) \propto \frac{[c_k(c_k + \beta_{k+1})\lambda_k\lambda_{k+1}]^u}{\Gamma(u+1)\Gamma(\alpha_{k+1}+u)}$

- $f(\theta | \lambda, data) \propto f(\theta) \exp\left\{\sum_{i=1}^{n_u} \theta' Z_i - \sum_{k=1}^{\infty} \lambda_k m_k(\theta)\right\}$

5

where $n_k = \sum_{i=1}^{n_u} I_{(\tau_{k-1}, \tau_k]}(t_i)$ and $m_k(\theta) = \sum_{i=1}^{n} W_{i,k}(t_i, \theta)$.

Similarly as we pointed out in the previous two cases, we incorporate a hyper prior process for the $\{c_k\}$ such that $c_k \sim Ga(1, \epsilon_k)$. The set of full conditional posterior distributions can then be extended to include

$$f(c_k|u_k, \lambda_k, \lambda_{k+1}) \propto (\beta_{k+1} + c_k)^{\alpha_{k+1}+u_k} c_k^{u_k} \exp\left\{-c_k\left(\lambda_{k+1} + \lambda_k + \frac{1}{\epsilon}\right)\right\}$$

Dependence between c_k's can be introduced through a hierarchical model via assigning a distribution to $\epsilon \sim Ga(a_0, b_0)$. So the update would be given by:

$$f(\epsilon|\{c_k\}) = Ga\left(\epsilon|a_0 + K, b_0 + \sum_{k=1}^{K} c_k\right)$$

where K is the number of intervals generated by the time axis partition. This hierarchical specification of the initial distribution for c_k let us assign a better value for c_k.

Gamma model examples

For this model, we will be using the control observations of the 6-MP data set (Freireich, E. J., et al., 1963) –data from a trial of 42 leukemia patients organised by pairs in which the first element of the pair is treated with a drug and the other is control–. For examples 1 to 4, we use a partition of unitary length intervals; for the last three examples –5, 6 & 7–, we use uniformly-dense intervals. The `"time"` column is taken as the observed times vector –`times`– and the `"cens"` column as the censoring status vector –`delta`–:

```
data(gehan)
times <- gehan$time[gehan$treat == "6-MP"]
delta <- gehan$cens[gehan$treat == "6-MP"]
```

Now that we have a time and censoring status vector, we can run several examples for this model. Default values are used for each function unless otherwise noted. Every example shows our estimate overlapped with the Nelson-Aalen / Kaplan-Meier estimator, so the user can compare them.

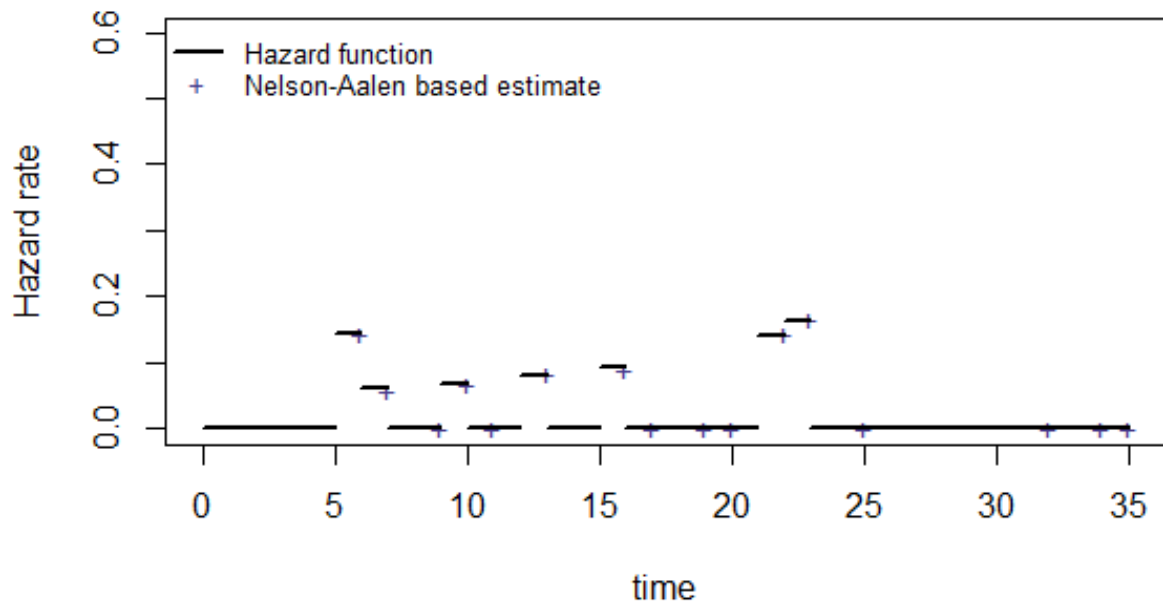### Example 1. Independence case. Unitary length intervals}

We obtain with our model the Nelson-Aalen and Kaplan-Meier estimators by defining c_k as a null vector –through fixing `type.c = 1`–. A unitary partitioned axis is obtained by fixing `type.t = 2`. In Figure 1 we show that our estimators –under independence– return the same results as the Nelson-Aalen and Kaplan-Meier estimators.

```
ExG1 <- GaMRes(times, delta, type.t = 2, K = 35, type.c = 1,
               iterations = 3000)
GaPloth(ExG1, confint = FALSE)
```

### Example 2. Introducing dependence through c. Unitary length intervals

The influence of c –or `c.r`– can be understood as a dependence parameter: the greater the value of each c_k, $k = 0, 1, 2, ..., K - 1$, the higher the dependence between intervals k and $k + 1$. For this example, we assign a fixed value to vector c, so we use `type.c = 2`. Comparing with the previous example, we see that the estimates that were zero, now have a positive value –see Figures 1 and 2–. Note how this model compares to Example 5 –see Figure 6–. The difference between those examples is how the partition is defined.

Estimate of hazard rates

(a) Hazard rates

Estimate of Survival Function
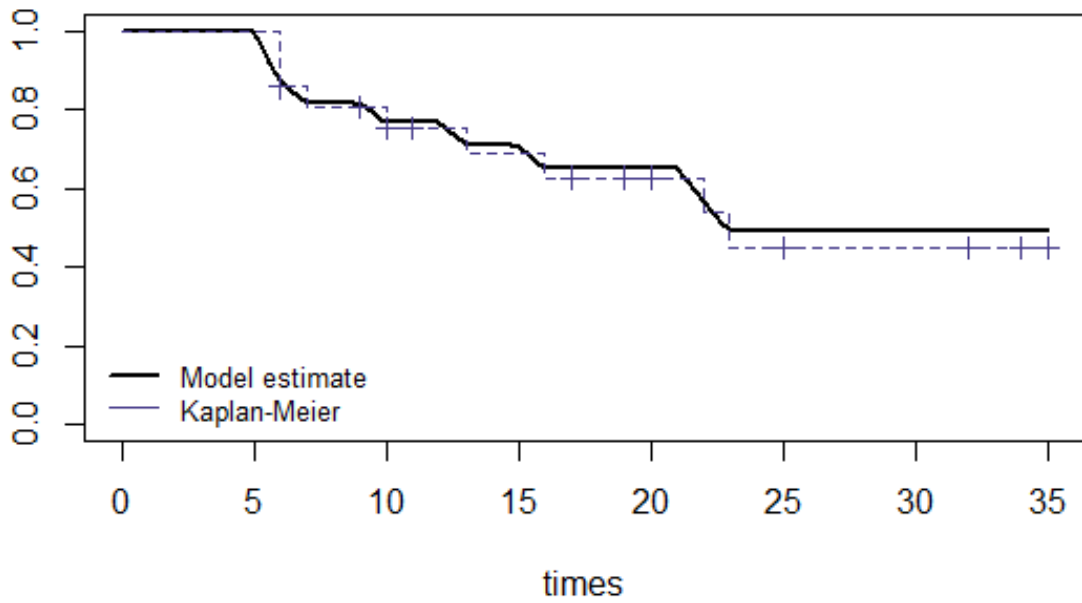
(b) Survival function

Figure 1: Gamma Example 1 - Independence case. Unitary length intervals

```
ExG2 <- GaMRes(times, delta, type.t = 2, K = 35, type.c = 2,
               c.r = rep(50, 34), iterations = 3000)
GaPloth(ExG2)
```

Additionally, we can get further detail on the Gibbs sampler with a diagnosis of the resulting Markov chain. We can run this diagnosis for each entry of λ, u, c or ϵ. In Figure 3 we show the diagnosis for λ_6 which includes the trace, the ergodic mean, the ACF function and the histogram for the generated chain.

```
GaPlotDiag(ExG2, variable = "lambda", pos = 6)
```

Example 3. Varying c through a distribution. Unitary length intervals

As we reviewed, we can learn about the $\{c_k\}$ via assigning an exponential distribution with mean ϵ. The estimates using $\epsilon = 1$ –type.c = 3– and unitary length intervals –type.t = 2– are shown in Figure 4. We can compare the hazard function with the previous example, where c was fixed, and observe that because of the variability given to c, the change on the estimated values between two contiguous intervals are greater in this example. The survival function echoes the shape of the Kaplan-Meier with higher decrease rates.

Comparing this example with Example 6 –Figure 7–, as with the previous example, the main difference is given by the partition of the time axis. This affects the estimates as we will note later.

```
ExG3 <- GaMRes(times, delta, type.t = 2, K = 35, type.c = 3, epsilon = 1,
               iterations = 3000)
GaPloth(ExG3)
```

Example 4. Using a hierarchical model to estimate c. Unitary intervals

Previous example can be extended with a hierarchical model, assigning a distribution to $\epsilon \sim Ga(a_0, b_0)$ with $a_0 = b_0 = 0.01$. In order to set up the model we should set type.c = 4. The result displayed on Figure 5 is a soft hazard function and a survival function that decreases faster than the Kaplan-Meier estimate.

```
ExG4 <- GaMRes(times, delta, type.t = 2, K = 35, type.c = 4,
               iterations = 3000)
GaPloth(ExG4)
```
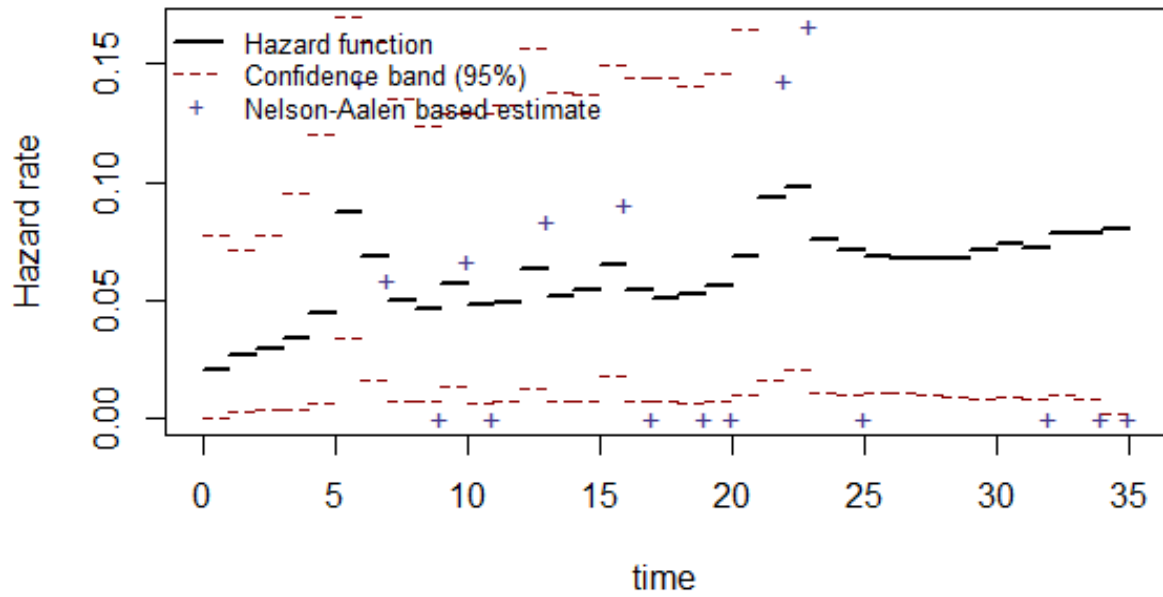
Example 5. Introducing dependence through c. Equally dense intervals

This example illustrates the same concept as Example 2 –how c introduces dependence–, but with a different partition of the time axis. To get this partition we set type.t = 1. Figure 6 shows that the survival function is close to the Kaplan-Meier estimate. The fact that it gets closer to the K-M estimates does not makes it a better estimate, we only could say that this partition yields, in average, a smaller hazard rate than the Example 2 built with a unitary length partition.

```
ExG5 <- GaMRes(times, delta, type.t = 1, K = 8, type.c = 2,
               c.r = rep(50, 7), iterations = 3000)
GaPloth(ExG5)
```
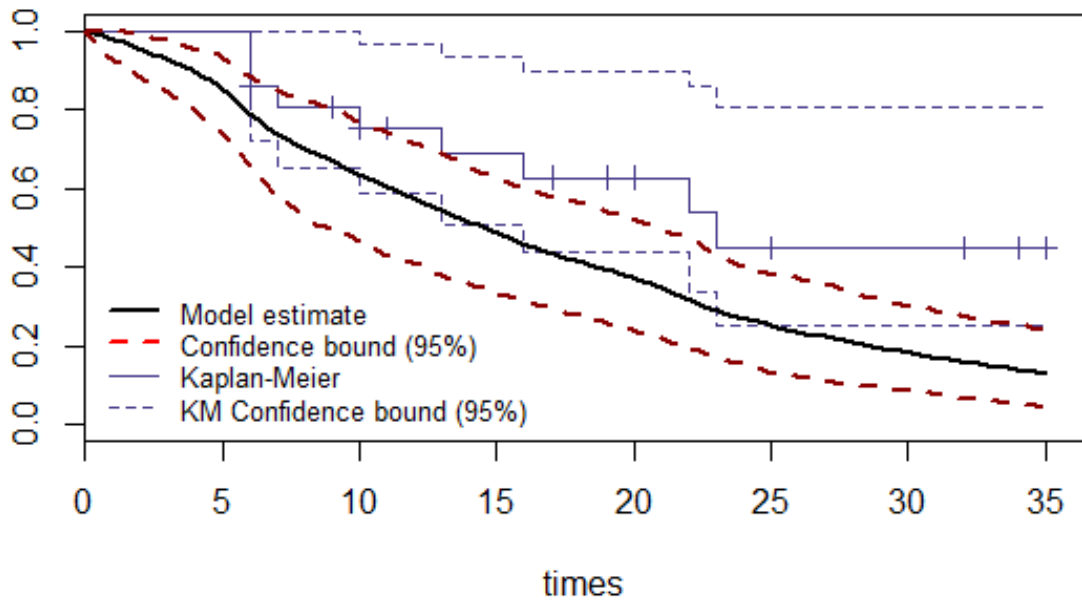
(a) Hazard rates

(b) Survival function

Figure 2: Gamma Example 2 - Introducing dependence through c ($c_k = 50, \forall k$). Unitary length intervals
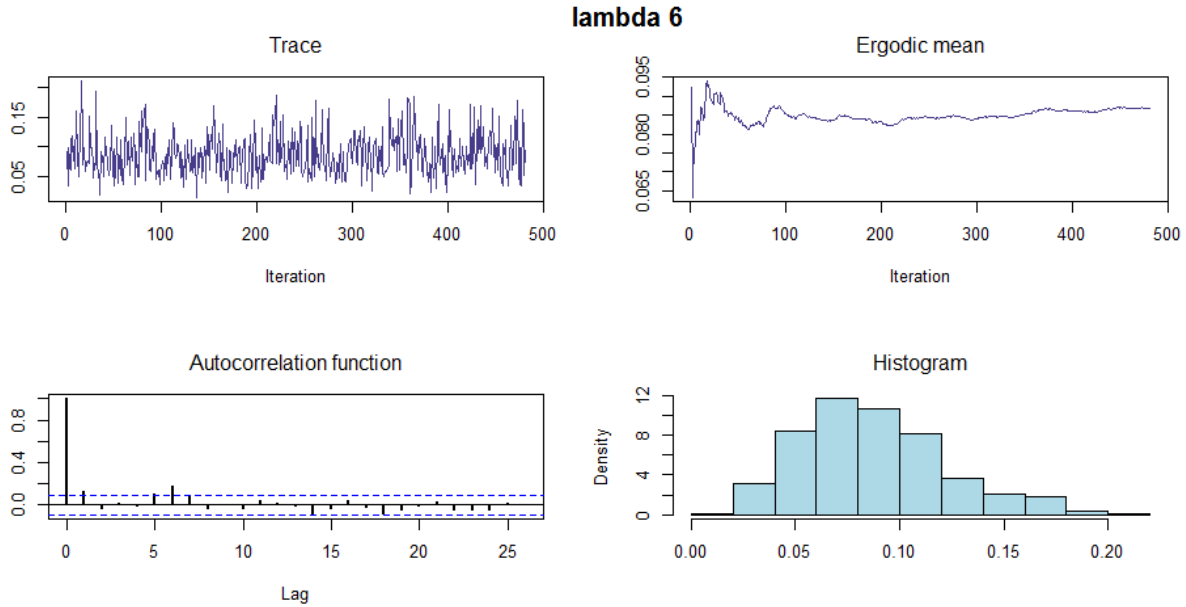
lambda 6

Figure 3: Gamma Example 2 - Diagnosis for λ_6

Example 6. Varying c through a distribution. Equally dense intervals

We can compare this example on Figure 7 with Example 3 –Figure 4–. We use less intervals, so the survival function is smoother. Our estimate decreases faster than the Kaplan-Meier estimate.

```
ExG6 <- GaMRes(times, delta, type.t = 1, K = 8, type.c = 3,
               iterations=3000)
GaPloth(ExG6)
```

Example 7. Using a hierarchical model to estimate c. Equally dense intervals

The survival curve for this particular example results in a smoothed version of the Kaplan-Meier estimate (Figure 8). As with previous examples, it can be compared with the unitary partition example (see Example 4, Figure 5).

```
ExG7 <- GaMRes(times, delta, type.t = 1, K = 8, type.c = 4,
               iterations = 3000)
GaPloth(ExG7)
```

%

10

Estimate of hazard rates
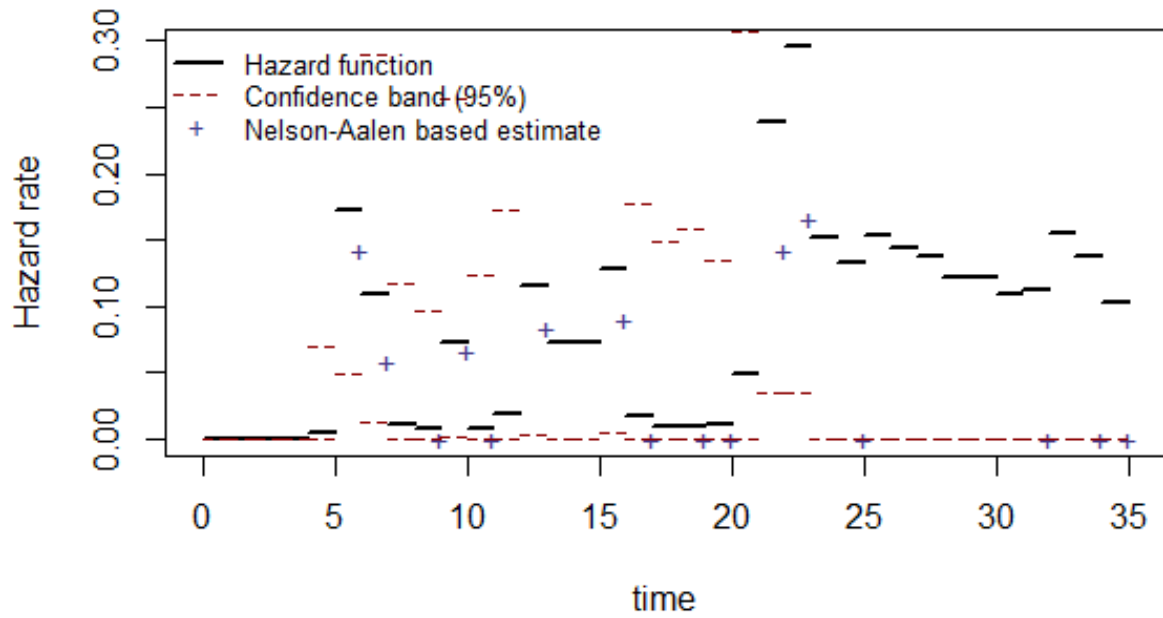
(a) Hazard rates

Estimate of Survival Function
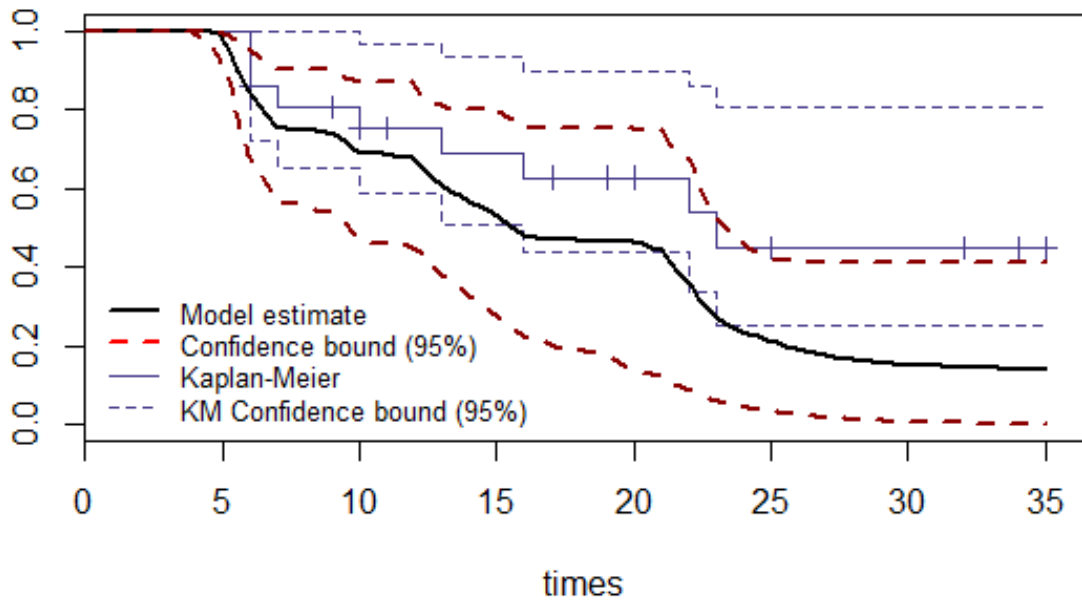
(b) Survival function

Figure 4: Gamma Example 3 - Varying c through a distribution $c_k \sim Ga(1, \epsilon = 1)$. Unitary length intervals
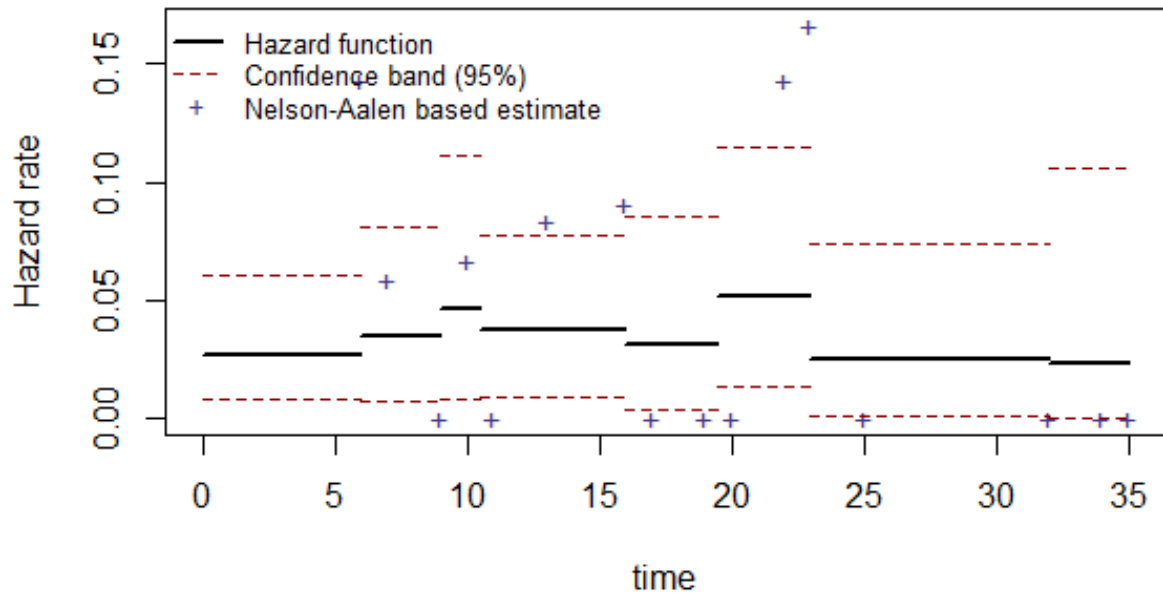
Estimate of hazard rates
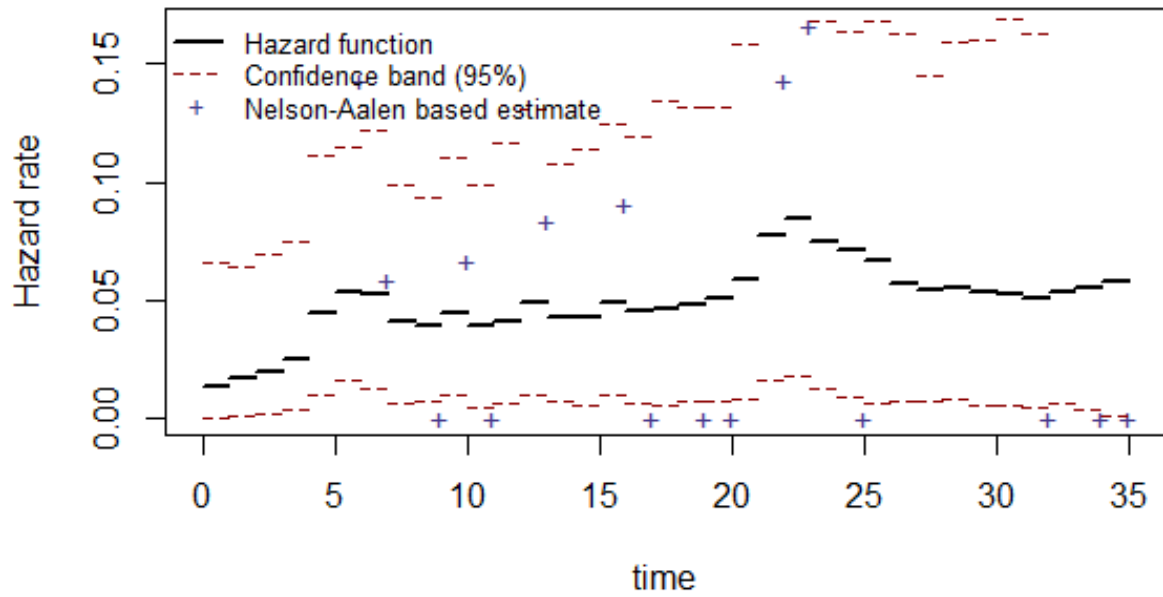
(a) Hazard rates
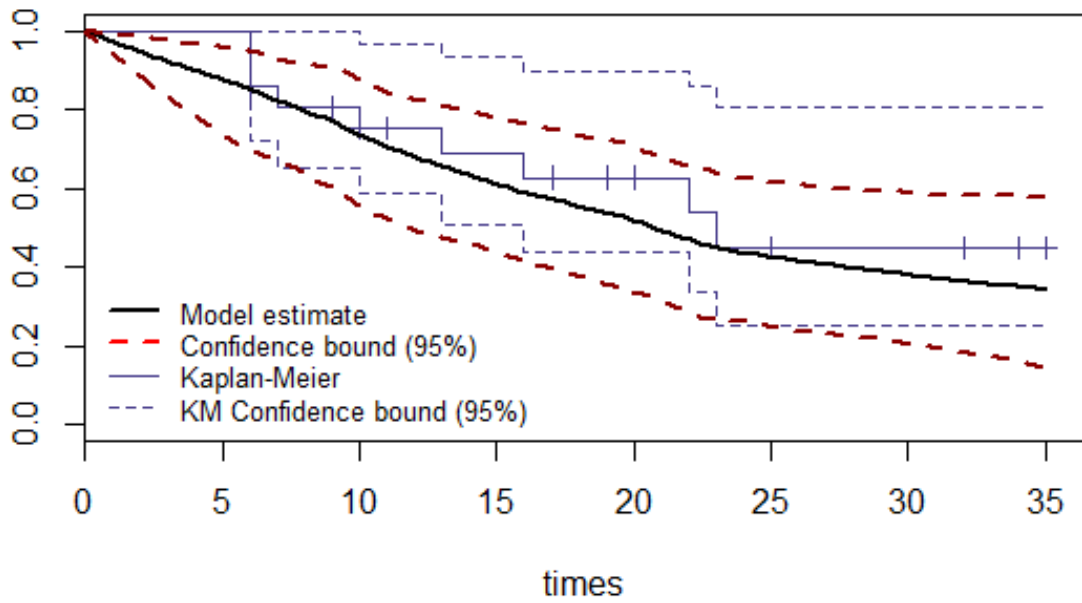
Estimate of Survival Function
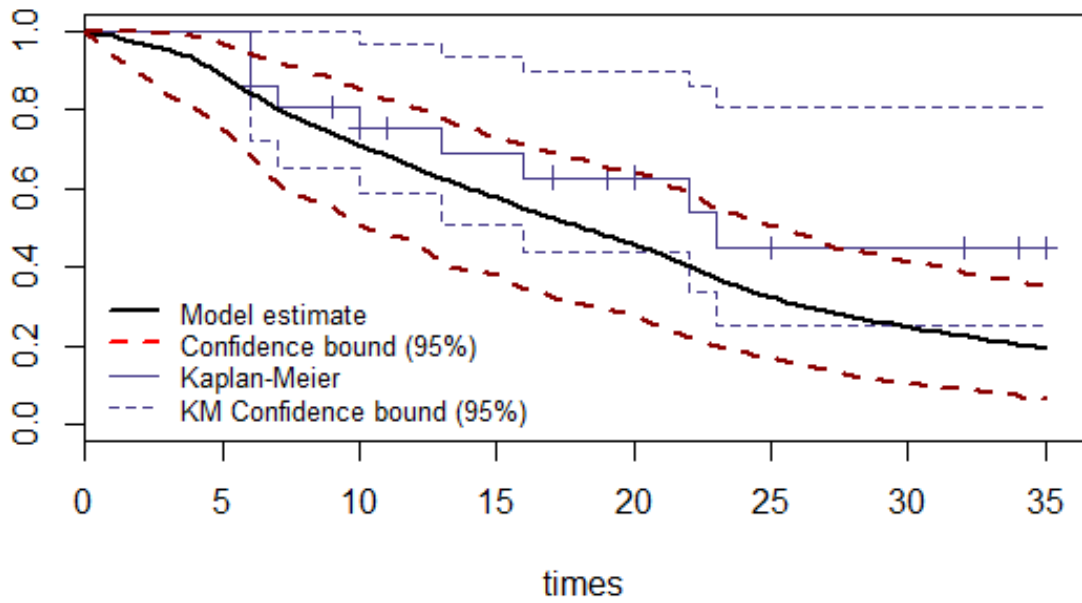
(b) Survival function

Figure 5: Gamma Example 4 - Using a hierarchical model to estimate c. Unitary intervals
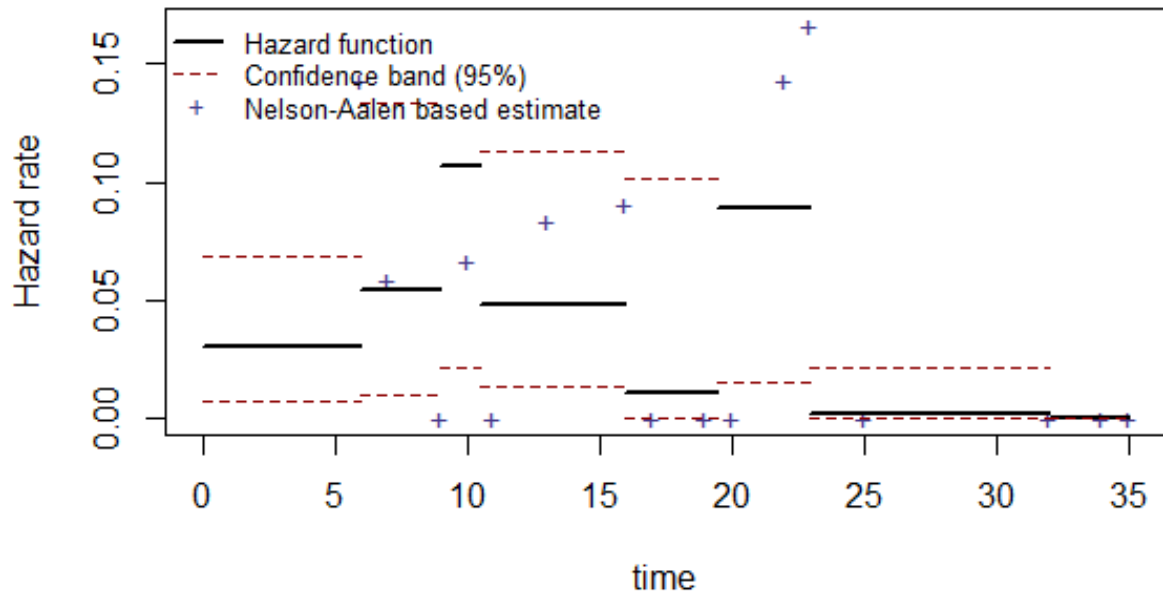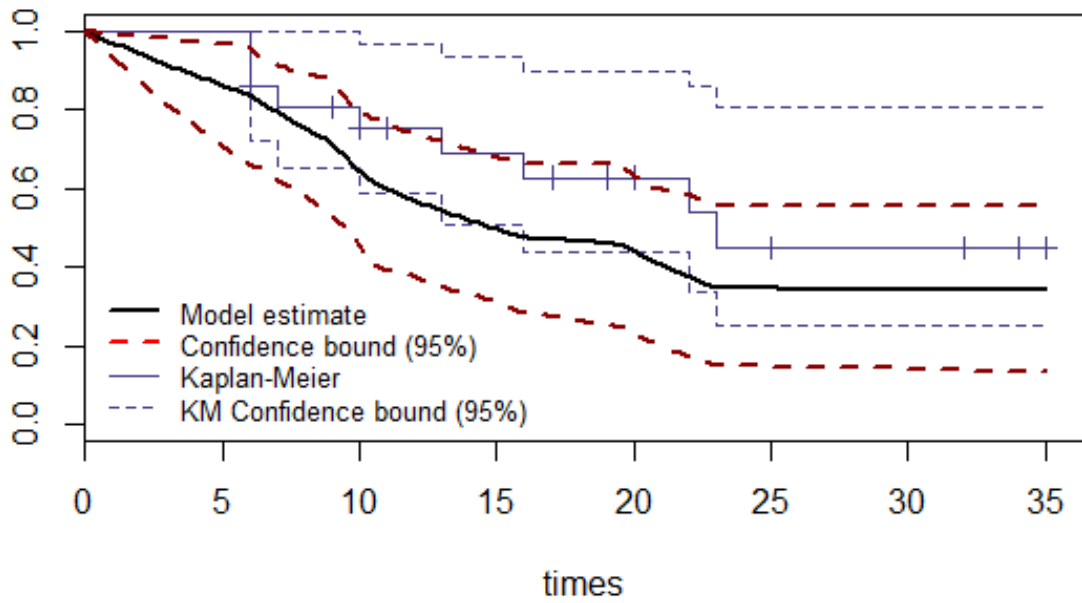
Estimate of hazard rates

(a) Hazard rates

Estimate of Survival Function

(b) Survival function

Figure 6: Gamma Example 5 - Introducing dependence through c ($c_k = 50, \forall k$). Equally dense intervals
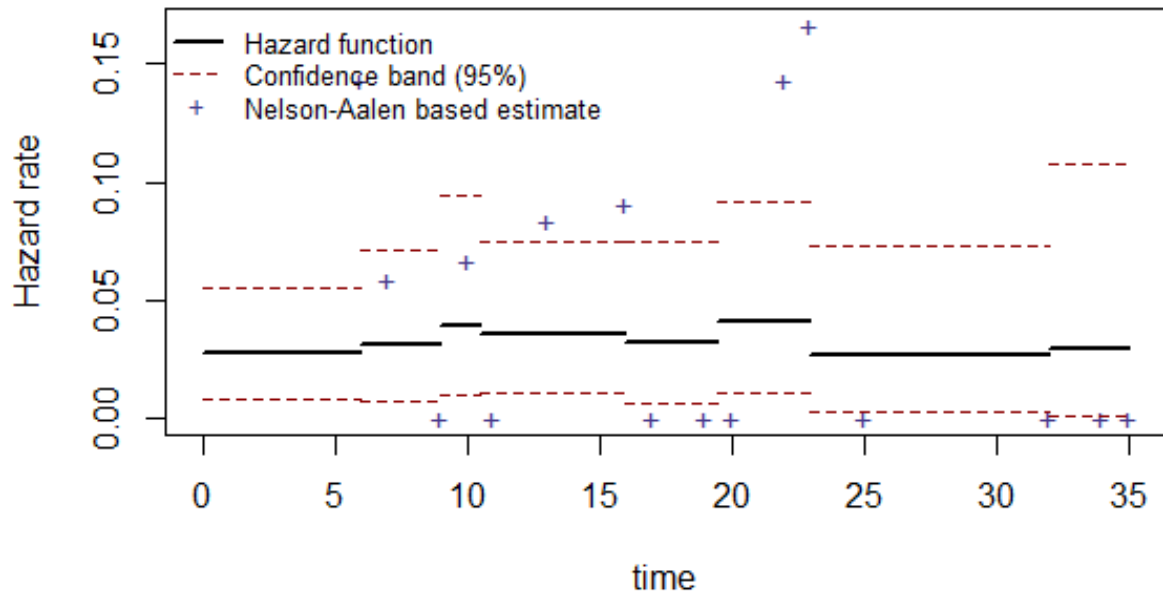
Estimate of hazard rates

(a) Hazard rates
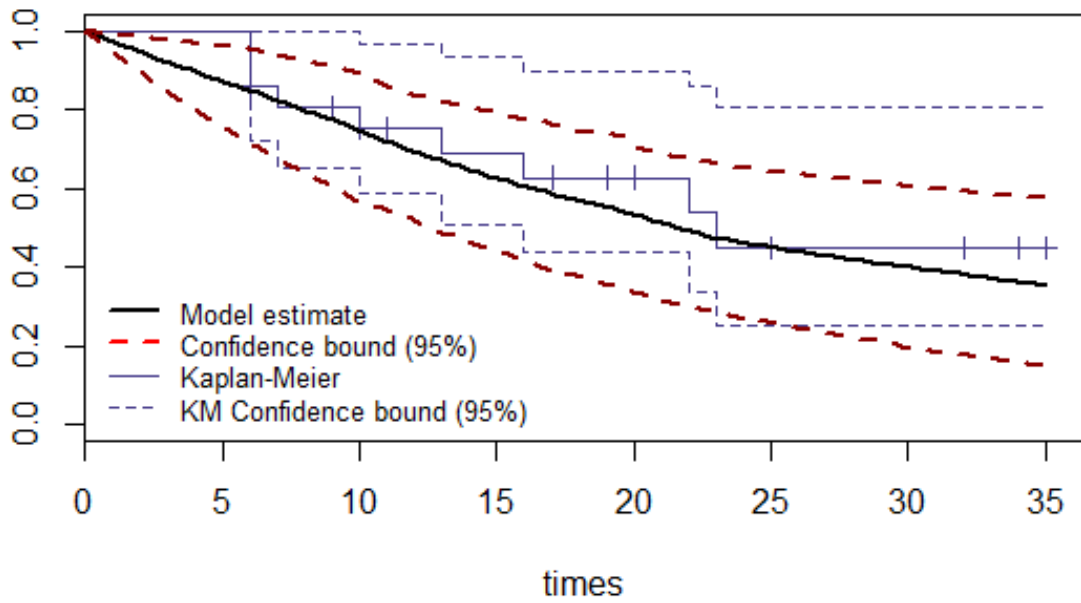
Estimate of Survival Function

(b) Survival function

Figure 7: Gamma Example 6 - Varying c through a distribution $c_k \sim Ga(1, \epsilon = 1)$. Equally dense intervals

14

Estimate of hazard rates

(a) Hazard rates

Estimate of Survival Function

(b) Survival function

Figure 8: Gamma Example 7 - Using a hierarchical model to estimate c. Equally dense intervals

15

Beta model examples

For this model, we use survival data on 26 psychiatric inpatients admitted to the University of Iowa hospitals during the years 1935-1948. This sample is part of a larger study of psychiatric inpatients discussed by Tsuang and Woolson (1977). We take the `"time"` column as the observed times vector –`times`– and the `"death"` column as the censoring status vector –`delta`–:

```
data(psych)
times <- psych$time
delta <- psych$death
```

Example 1. Independence case

As with the Gamma Example, we obtain the Nelson-Aalen and Kaplan-Meier estimators by defining c_k as a null vector through fixing `type.c = 1` –see Figure 9– The conclusion does not change: the independence case of our model results on the N-A and K-M estimators.

```
ExB1 <- BeMRes(times, delta, type.c = 1, iterations = 3000)
BePloth(ExB1, confint = FALSE)
```

Example 2. Introducing dependence through c

The influence of c –or `c.r`– can be also understood as a dependence parameter: the greater the value of each c_k, $k = 0, 1, 2, ..., K-1$, the higher the dependence between intervals k and $k+1$. In this example, we fix each c entry at 100. As we are defining vector c with fixed values, we should fix `type.c = 2`. We see on Figure 10 that the hazard function estimate turns smoother than on the previous example. The steps on the survival function appear to be more uniform than on the Kaplan-Meier estimate.

```
ExB2 <- BeMRes(times, delta, type.c = 2, c.r = rep(100, 39),
               iterations = 3000)
BePloth(ExB2)
```

Additionally, we can get further detail on the Gibbs sampler with a diagnosis of the resulting Markov chain. We can run this diagnosis for each entry from π, u, c or ϵ. In Figure 11 we show the diagnosis for π_{10} which includes plot of the trace, the ergodic mean, the ACF function and the histogram of the chain.

```
BePlotDiag(ExB2, variable = "Pi", pos = 6)
```
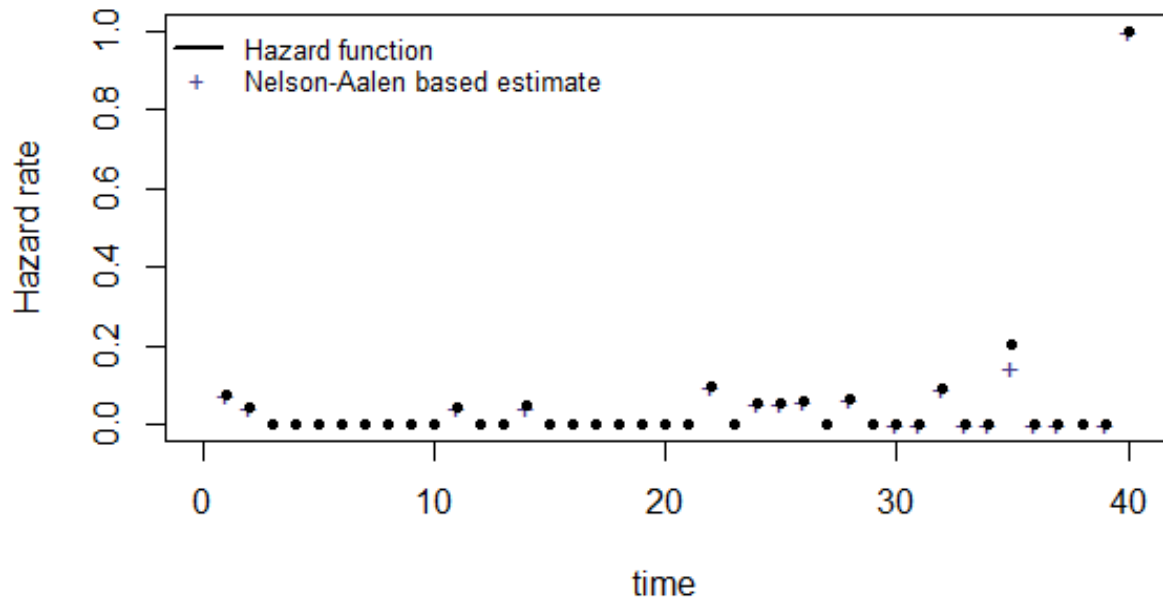
Example 3. Varying c through a distribution

As with the gamma model, we can learn about the $\{c_k\}$ via assigning an exponential distribution with mean ϵ. The estimates using $\epsilon = 1$ and unitary length intervals are shown on Figure 12. Note that the confidence intervals have widen: it is a signal that we have introduced variability to the model –because of the distribution assigned to c–.

```
ExB3 <- BeMRes(times, delta, type.c = 3, epsilon = 1, iterations = 3000)
BePloth(ExB3)
```
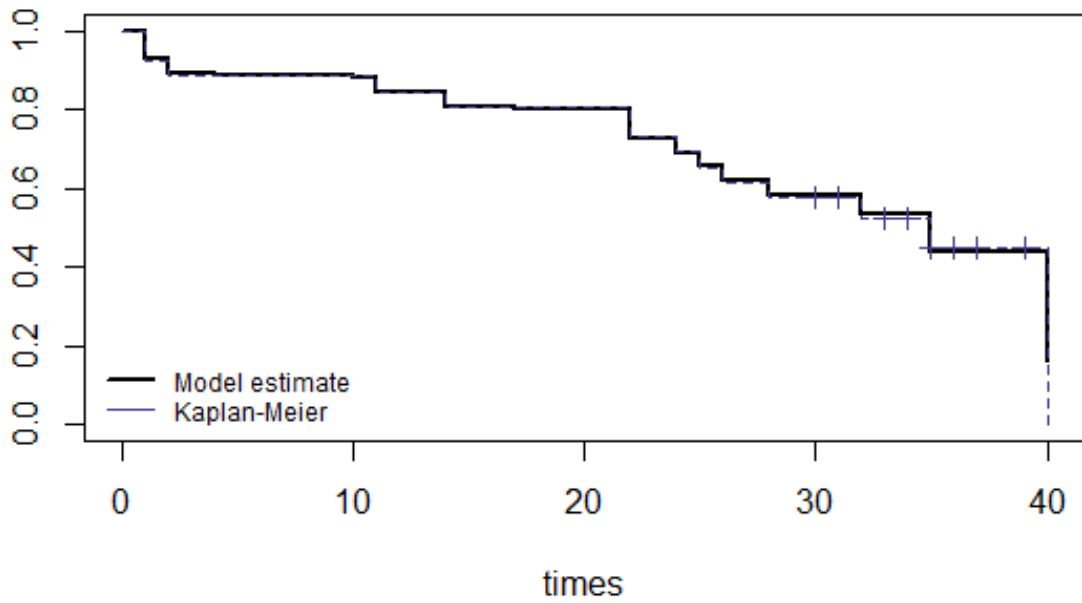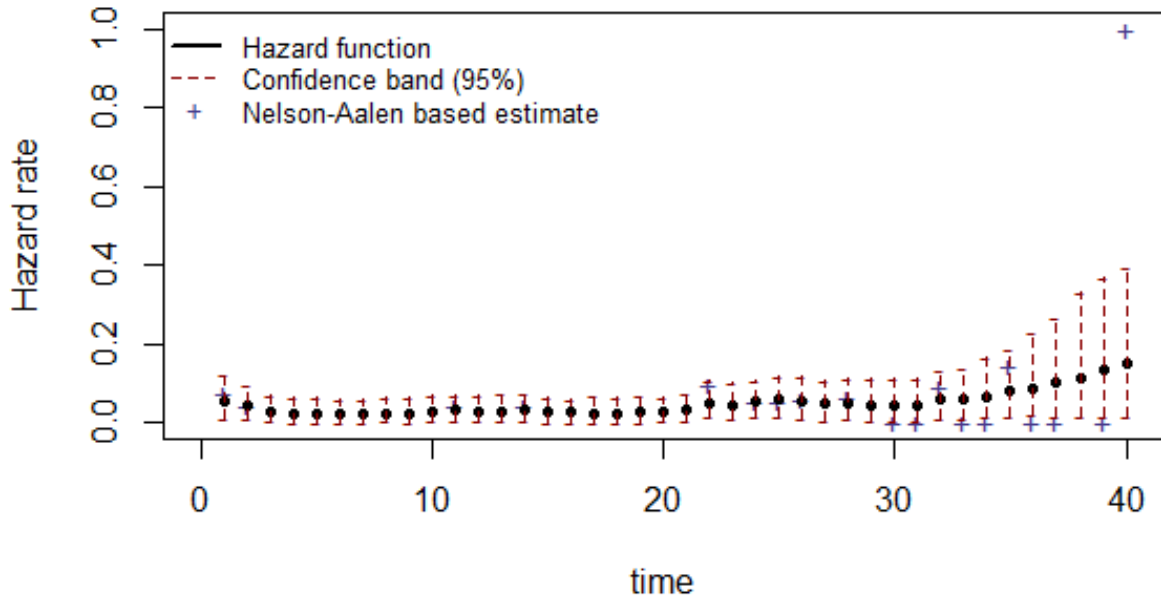
Estimate of hazard rates

Hazard function
Nelson-Aalen based estimate

(a) Hazard rates

Estimate of Survival Function

Model estimate
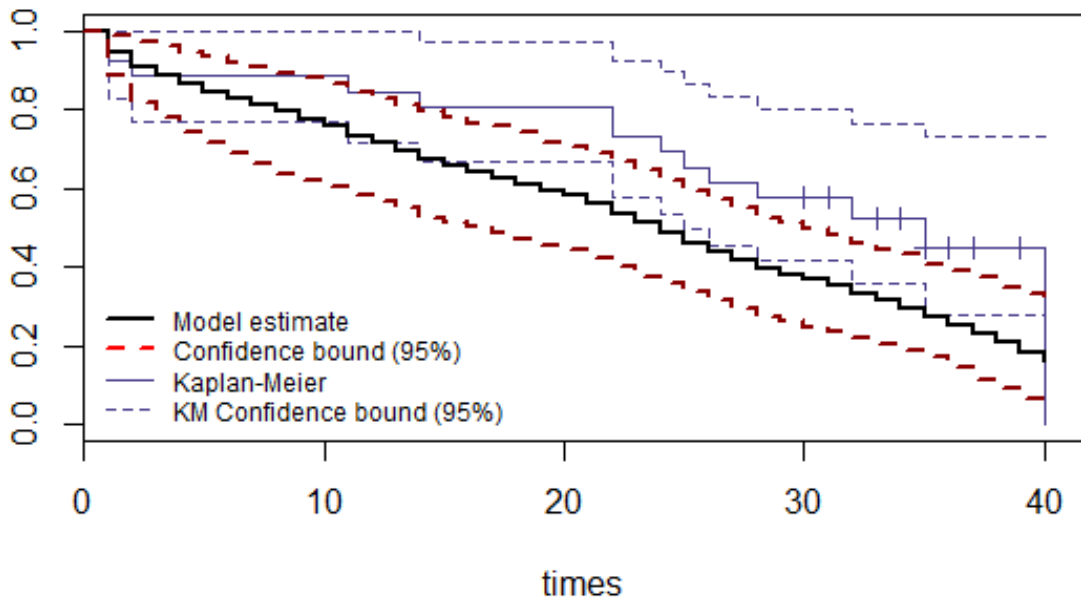Kaplan-Meier

(b) Survival function

Figure 9: Beta Example 1 - Independence case

17

Estimate of hazard rates

(a) Hazard rates

Estimate of Survival Function

(b) Survival function

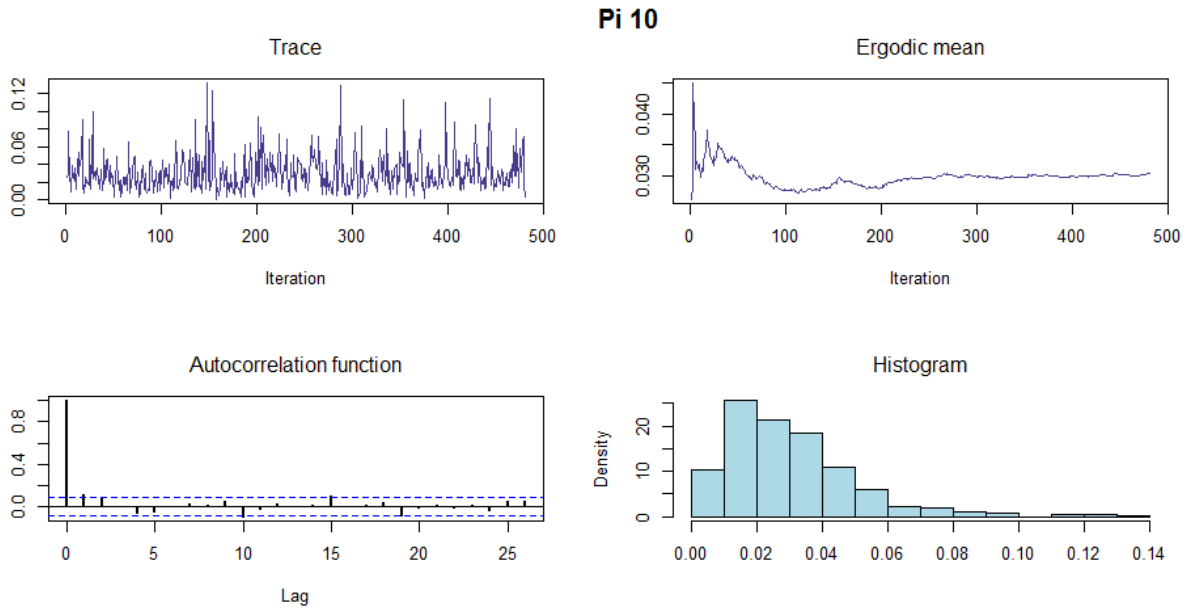Figure 10: Beta Example 2 - Introducing dependence through c ($c_k = 100, \forall k$)

Pi 10

Figure 11: Beta Example 2 - Diagnosis for π_{10}

Example 4. Using a hierarchical model to estimate c

The previous example can be extended with a hierarchical model, assigning a distribution to $\epsilon \sim Ga(a_0, b_0)$, with $a_0 = b_0 = 0.01$. In order to set up the model we should set `type.c = 4`. The result displayed on Figure 13 is a soft hazard function and a survival function that decreases faster than the Kaplan-Meier estimate.
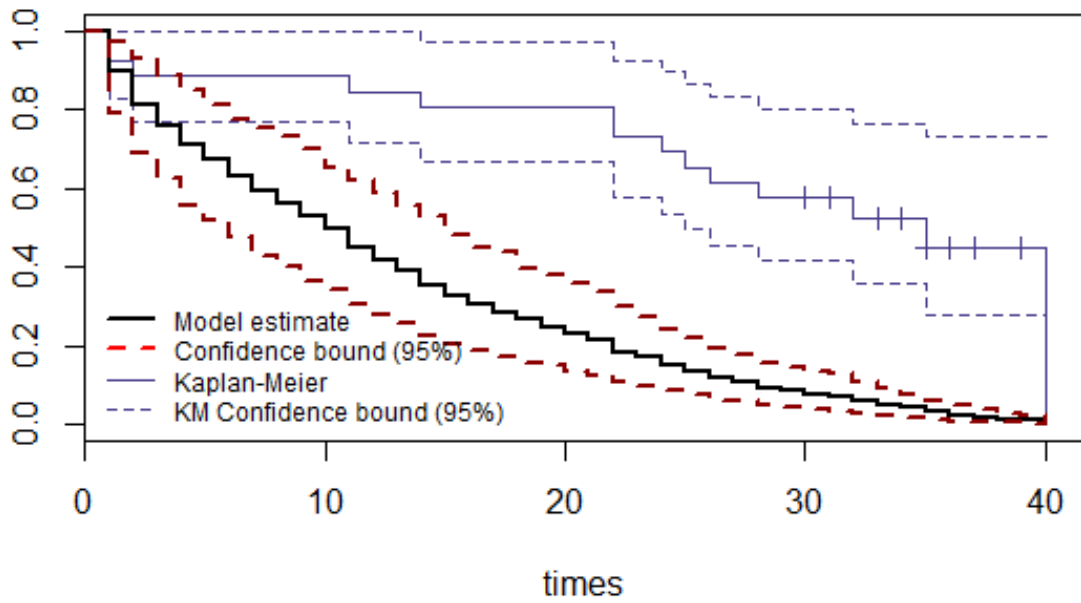
```
ExB4 <- BeMRes(times, delta, type.c = 4, iterations = 3000)
BePloth(ExB4)
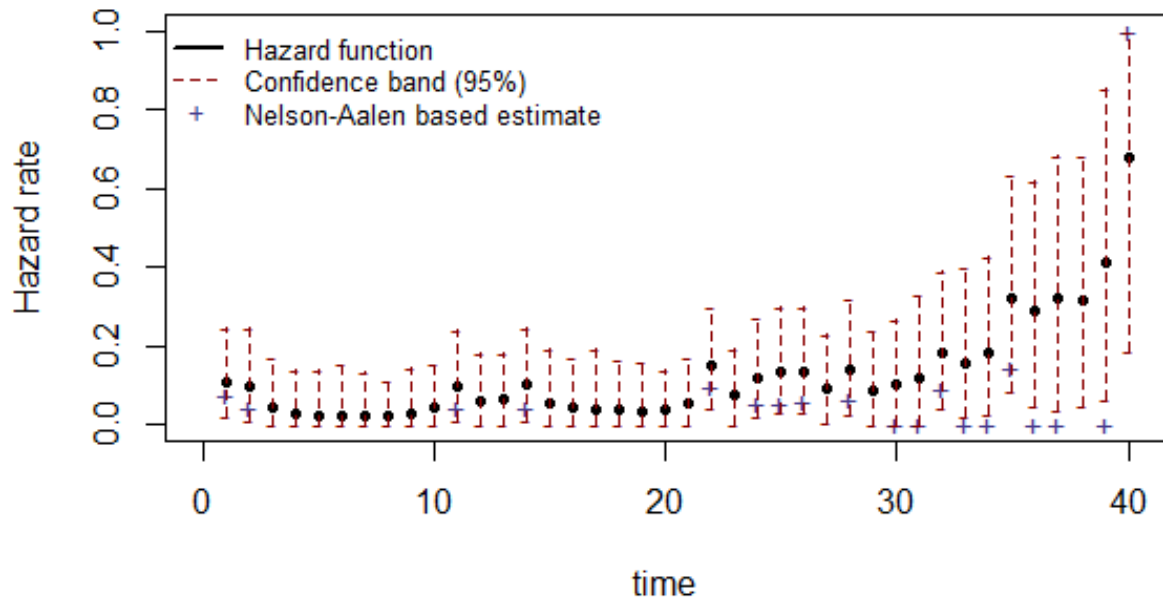```

19

## Estimate of hazard rates
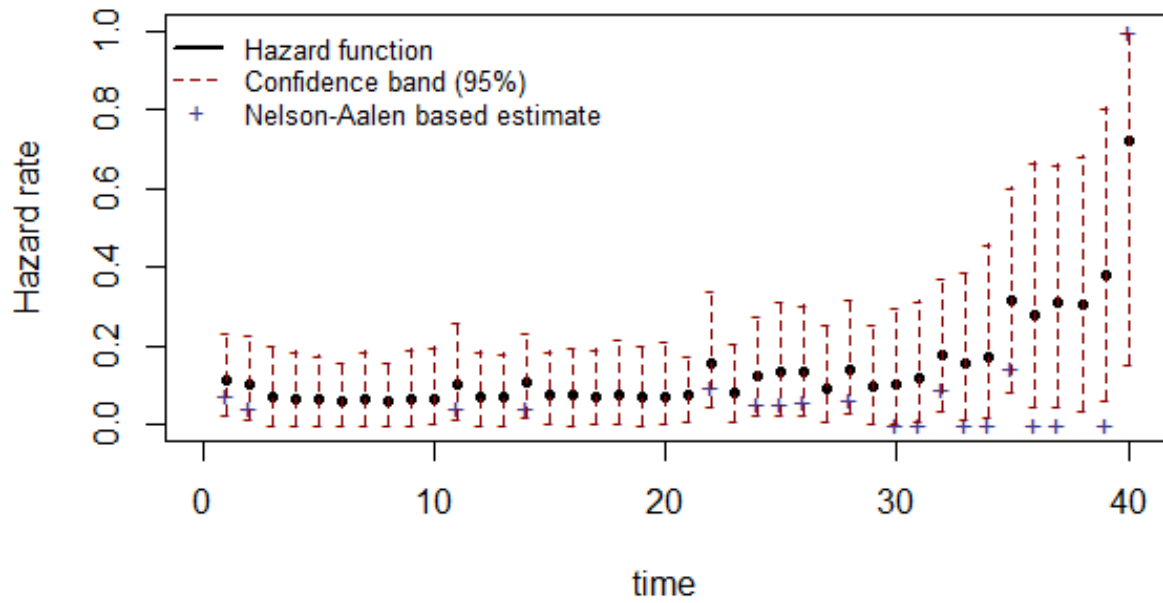


(a) Hazard rates

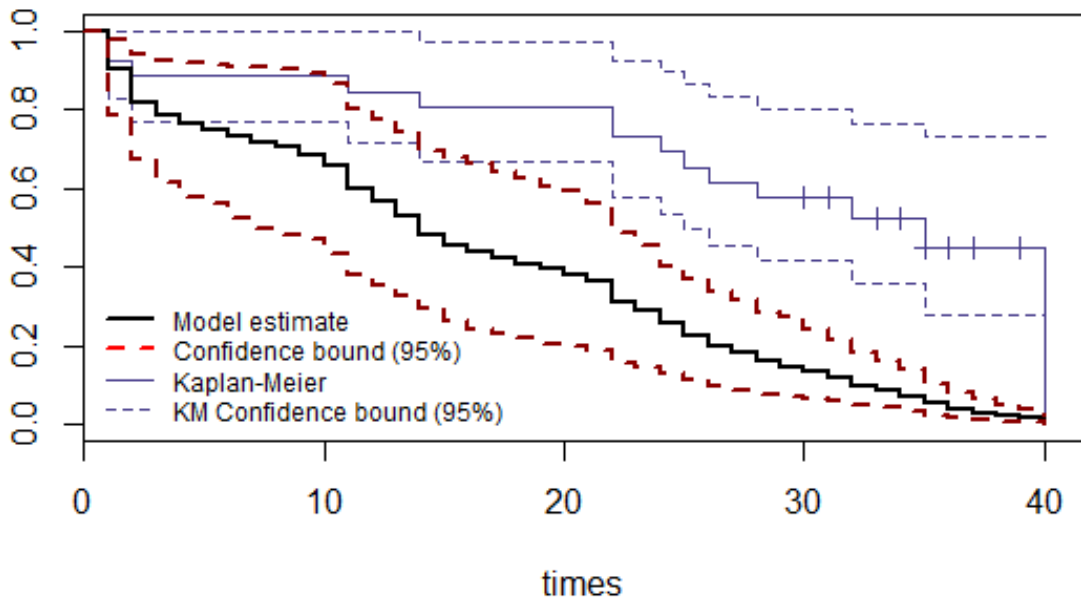## Estimate of Survival Function



(b) Survival function

Figure 12: Beta Example 3 - Varying $c$ through a distribution $c_k \sim Ga(1, \epsilon = 1)$

## Estimate of hazard rates



(a) Hazard rates

## Estimate of Survival Function



(b) Survival function

Figure 13: Beta Example 4 - Using a hierarchical model to estimate $c$

| i | $t_i$ | $Z_{i1} \sim U(0,1)$ | $Z_{i2} \sim U(0,1)$ | $c_i \sim Exp(1)$ | $\delta_i = I(t_i > c_i)$ | $min\{t_i, c_i\}$ |
|---|-------|----------------------|----------------------|-------------------|---------------------------|-------------------|
| 1 | $t_1$ | $Z_{11}$ | $Z_{12}$ | $c_1$ | $\delta_1$ | $min\{t_1, c_1\}$ |
| 2 | $t_2$ | $Z_{21}$ | $Z_{22}$ | $c_2$ | $\delta_2$ | $min\{t_2, c_2\}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $n$ | $t_n$ | $Z_{n1}$ | $Z_{n2}$ | $c_n$ | $\delta_n$ | $min\{t_n, c_n\}$ |

Table 1: Weibull simulation model

## Cox-gamma model example

In this example, we simulate the data from a Weibull model, frequently used with continuous and non negative data. The advantage of setting a model is that we know in advance the results, so we can compare the estimates with the exact values from model. The Weibull model has the probability functions:

$$h_0(t) = abt^{b-1}, S_0(t) = e^{-at^b}$$

We construct the proportional hazard model as:

$$h_i(t_i|Z_i) = h_0(t)e^{\theta' Z_i}bt^{b-1}, S_i(t_i|Z_i) = exp\left\{-H_0(t_i)e^{\theta' Z_i}\right\}$$

Note that this Weibull proportional hazard model is also another Weibull model with parameters ($a_i^* = ae^{\theta' Z_i}, b$). Based on the previous densities and on fixed parameters $a, b, \theta_1$ and $\theta_2$ -these last two covariates are simulated from uniform distributions on the interval (0,1)-, we simulate $n$ observations. We gather the results of the model in a table –see Table 1–. $t_i|Z_i \sim Weibull(a_i^*, b)$ and $Z_i = (Z_{i1}, Z_{i2})$ are the explanatory variables; $c_i$, the censoring time; $\delta_i$, censoring indicator, and $min\{t_i, c_i\}$ represents *observed* values deeming censorship time.

We generate a size $n = 100$ sample based on the simulation model with parameters $a=$0.1, $b = 1, \theta = (1, 1)$ y $Z_i \sim U(0,1)$, $i = 1, 2$. The result is a table with $n = 100$ observations.

On the other hand, we use almost every default parameter from `CGaMRes` excluding K = 10, `iterations = 3000` and `thpar = 10`. Theoretically, our model should estimate a constant risk function at $a \times b = 0.1$.

Below, we show the code for the Weibull model and the calls for the plots for the hazard and survival functions –command `CGaPloth(M)`–, the predictive distribution for an observation defined as the median of the data –`CGaPred(M)`–, and the plots for $\theta_1$ and $\theta_2$ –`PlotTheta(M)`–.

```
SampWeibull <- function(n, a = 10, b = 1, beta = c(1, 1)) {
  M <- matrix(0, ncol = 7, nrow = n)
  for(i in 1:n){
    M[i, 1] <- i
    M[i, 2] <- x1 <- runif(1)
    M[i, 3] <- x2 <- runif(1)
    M[i, 4] <- rweibull(1, shape = b,
                    scale = 1 / (a * exp(cbind(x1, x2) %*% beta)))
    M[i, 5] <- rexp(1)
    M[i, 6] <- M[i, 4] > M[i, 5]
    M[i, 7] <- min(M[i, 4], M[i, 5])
  }
  colnames(M) <- c("i", "x_i1", "x_i2", "t_i", "c_i", "delta",
                "min{c_i, d_i}")
  return(M)
}
```

```
dat <- SampWeibull(100, 0.1, 1, c(1, 1))
dat <- cbind(dat[, c(4, 6)], dat[, c(2, 3)])
CG <- CGaMRes(dat, K = 10, iterations = 3000, thpar = 10)
CGaPloth(CG)
PlotTheta(CG)
CGaPred(CG)
```
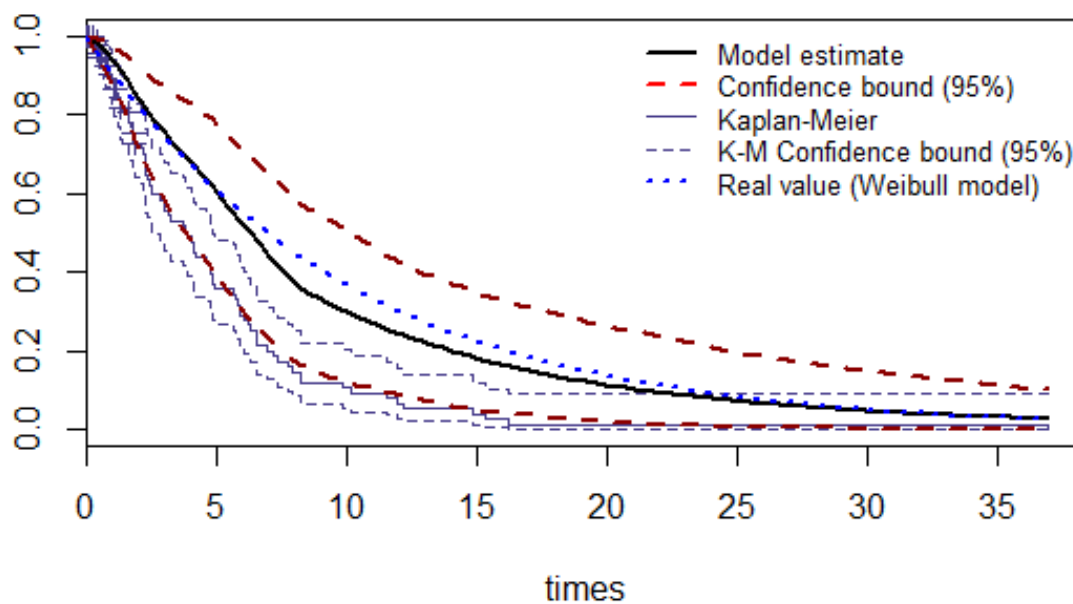
Because of the way we built the model, we can compare against theoretical results on the Weibull model. In Figure 14 we see that the hazard rate estimate is practically the same from t = 8 to t = 37. The estimate for the survival functions gets very close to the real value. Plots and histograms for θ from Figure 15 show estimated values for the regression coefficients (θ_1, θ_2) and they are consistently near to 1. Finally, the plots from Figure 16 show the hazard rate estimate over a equally dense partition for a future individual where its explanatory variable is equal to the median of the observations $-x_F-$. Note that the effect over the hazard function is given by the product of the baseline hazard function and $\exp\{x'_F\theta\}$.
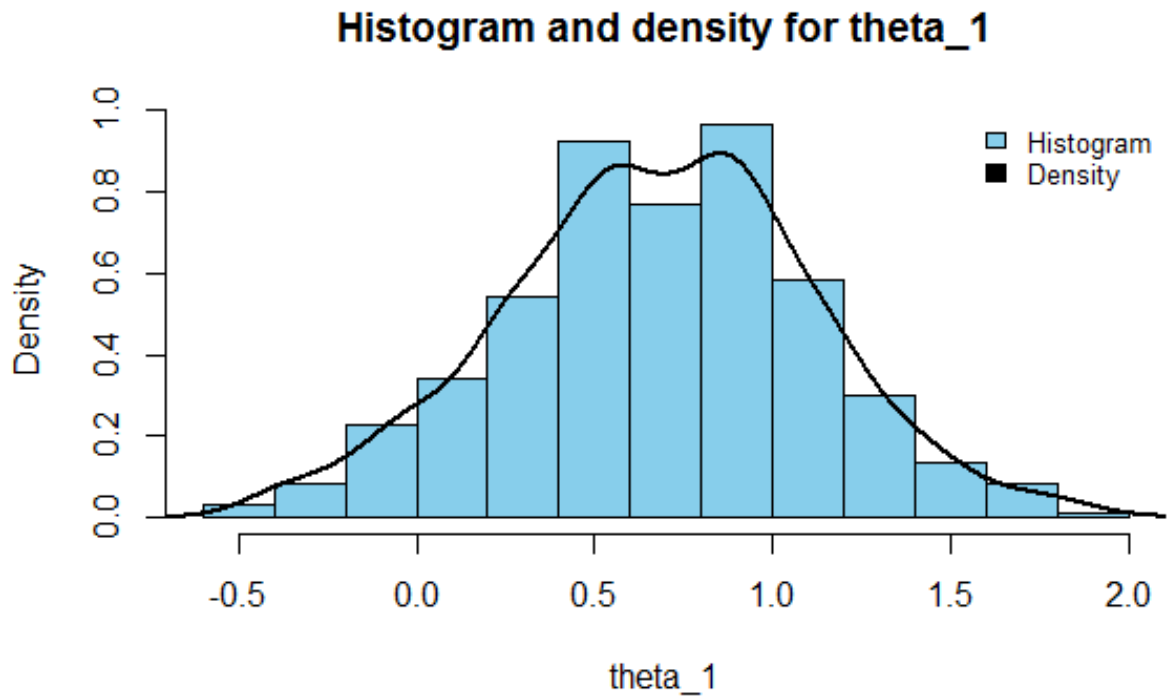
Baseline hazard estimate
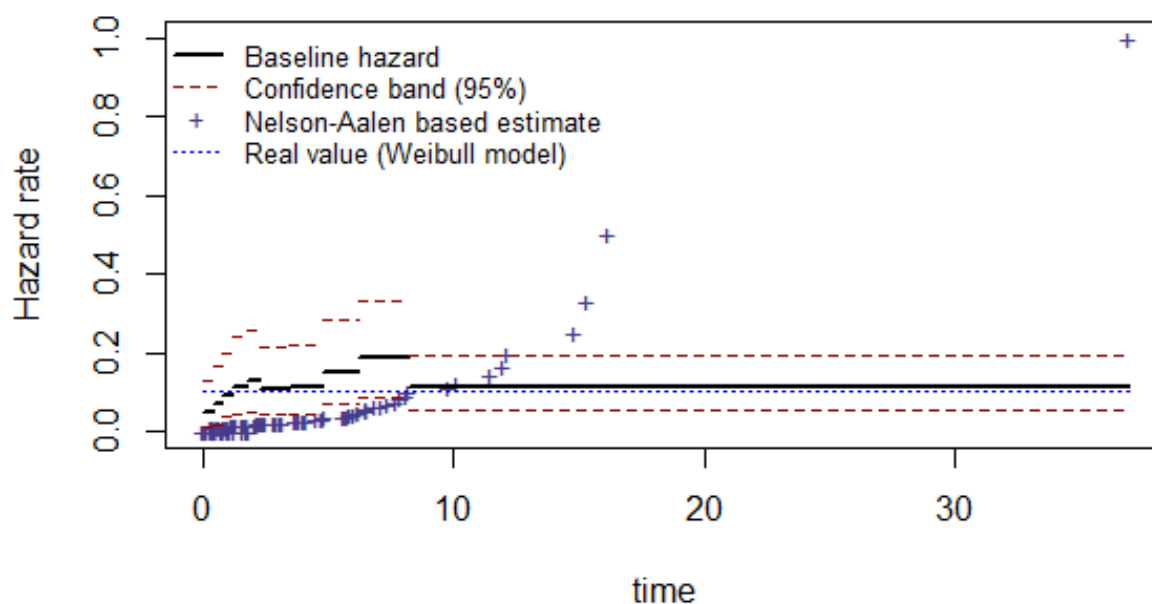
(a) Hazard rate estimate

Baseline survival function
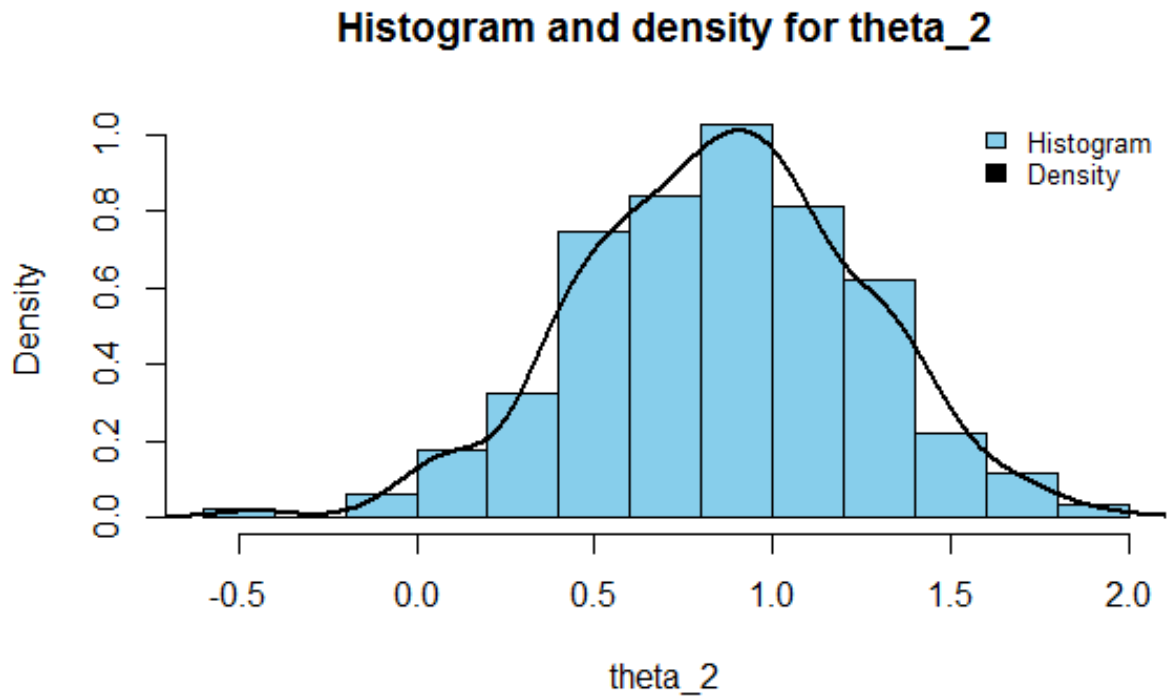

(b) Survival function estimate

Figure 14: Cox-gamma example

Histogram and density for theta_1

(a) θ_1 estimate

Histogram and density for theta_2

(b) θ_2 estimate

Figure 15: θ estimate on the Cox-gamma example
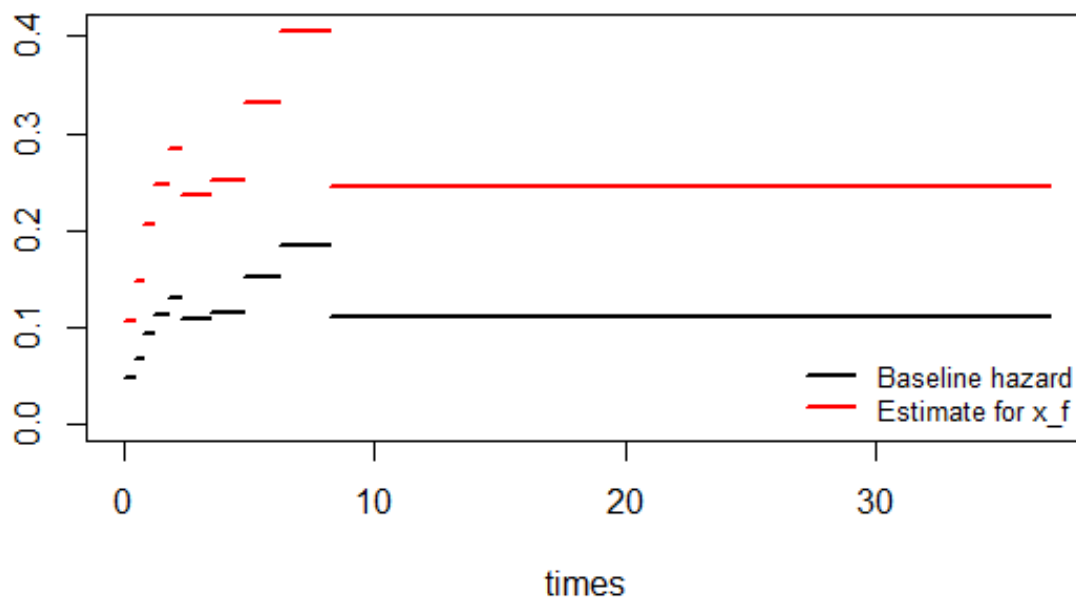
Hazard distribution estimate

Figure 16: Hazard rate estimate for the median on the Cox-gamma example

References

- FREIREICH. E.J., ET AL., Estimation of Exponential Survival Probabilities with Concomitant Information, *Biometrics*, **21**, pages: 826-838, 1965.

- GEHAN, E.A., A generalized Wilcoxon test for comparing arbitrarily single-censored samples, *Biometrika*, **52**, pages: 687-696, 1965.

- HJORT, N.L., Nonparametric Bayes estimators based on beta processes in models for life history data, *Annals of Statistics*, **18**, p'ags: 1259-1294, 1990.

- KAPLAN, E.L. y MEIER, P. Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* **53**. 282, p'ags: 457-481, 1958.

- KLEIN. J.P. y MOESCHBERGER, M.L. Survival analysis: techniques for censored and truncated data, Springer Science & Business Media, 2003.

- NIETO-BARAJAS, L.E. & WALKER, S.G., Markov beta and gamma processes for modelling hazard rates, *Scandinavian Journal of Statistics* no. **29**, pages 413-424, 2002.

- NIETO-BARAJAS, L.E., Discrete time Markov gamma processes and time dependent covariates in survival analysis, *Bulletin of the International Statistical Institute 54th Session*, 2003.

- TSUANG, M.T. AND WOOLSON, R.F. Mortality in Patients with Schizophrenia, Mania and Depression, *British Journal of Psychiatry*, **130**, pages 162-166, 1977.

- WALKER, S.G. y MALLICK, B.K., Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing, *Journal of the Royal Statistical Society*, Series B **59**, p'ags: 845-860, 1997.

- WOOLSON, R.F., Rank Tests and a One-Sample Log Rank Test for Comparing Observed Survival Data to a Standard Population, *Biometrics*, **37**, pages: 687-696, 1981.