The importance of selecting Bayesian nonparametric survival models: application to the estimation of unemployment durations

🆔 HAMIMES Ahmed

ahmed.hamimes@univ-constantine3.dz

Faculty of medecine, University of Constantine 3, Algeria

Abstract:

The Kaplan-Meier and Fleming-Harrington estimator in the frequentist approach are functional methods. We compare in this article the different Bayesian structures of the Kaplan-Meier and Fleming-Harrington estimators through the information deviance criterion in a real example describes the duration of unemployment of 1064 individuals in an employment agency. employment at the local employment agency in Ain El Benian. This study clearly shows the importance of bayesian model selection in duration models.

Keywords: Fleming-Harrington, Kaplan-Meier, the Bayesian paradigm, unemployment durations.

Jel code: C41, C11, E24.

Introduction :

Users of survival analysis methods generally need to select an approach from the full panel available to them. This choice will depend not only on the form of survival to be measured, but also on the statistical approach to inference used. Several Bayesian analyzes have been performed using the Cox model (Ferguson (1973), Kalbfleisch (1978), Kalbfleisch and Prentice (1980) and Clayton (1978)), in which the prior covariates for the base rate and the coefficients of the regressions are defined in various ways. Both of these Bayesian approaches used full likelihood rather than partial likelihood. Indeed, one of the advantages of using Bayesian approaches to jointly model the covariate regression coefficients and the baseline mortality rate is that using MCMC techniques one can accurately measure the posterior distributions of the model and their standard deviations. However, the question of specification functions of rational prior distribution as well as performing intensive calculations remains. The results of these Bayesian proportional survival studies demonstrated the accuracy of the estimates and the possible benefits of using these approaches to assess survival data. We can cite the Bayesian approach to breakpoint models introduced by Carlin, Gelfand and Smith in parametric modeling in 1992. In nonparametric modeling in 2001, Florens and Rolin mainly demonstrated that estimates of the Dirichlet process by simulation and nonparametric Bayesian inference perform well over conventional methods. These different works have been constructed in various methodological contexts using various prior distributions and/or by modeling the cumulative risk function, or directly the instantaneous risk function.

Nonparametric modeling is not always pragmatic compared to parametric modeling because it aims to estimate an infinite number of parameters by a finite number of observations, but in the majority of duration models our sample is not large enough Due to the complexity of observed phenomena and resource constraint, also nonparametric Bayesian modeling requires greater theoretical background. In the nonparametric Bayesian approach, the modeling is based on a random measurement. In contrast, the Kaplan-Meier and Fleming-Harrington estimator in the frequentist approach are functional methods of estimating the survival function and take into account the presence of legitimate censoring. Moreover, the estimator, for example Kaplan-Meier, is convergent, consistent and asymptotically Gaussian and it is biased positively. Several works have been based on the improvement of these estimators, in the Kaplan Meier estimator we find: Khizanov and Maĭboroda (2015), proposed a modification based on a mixture model with varied concentrations, Rossa and Zieliński (2006) , introduce a Kaplan-Meier estimator based on an approximation by Weibull's

law, Shafiq Mohammad et al (2007), presented a weighting of the Kaplan Meier estimator under the sine function.

1. Nonparametric estimation methods

1.1. The Kaplan-Meier estimator

If $(\mathfrak{X}, \mathcal{F}, P)$ a probabilistic space. If $T_1, \dots, T_m$ a sequence of random variables i.i.d positive and with a common distribution function F, express the time between the start of the study and the arrival of an event for the ith individual The sample actually observed will therefore be composed of m-couples (Y_i, δ_i), where δ_i is the censorship flag, which determines whether T has been censored or not. Kaplan and Meier (1958) introduced a non-parametric estimator written in two different ways depending on whether we are in the absence of exo aequo or in the presence of exo aequo as follows :

- Case of no exo aequo:

The Kaplan-Meier estimator is given by:

$$\widehat{S}(t) = \begin{cases} \Pi_{T_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right)^{\delta_i} & \text{si } t \geq t_1 \\ 1 & \text{si } t < t_1 \end{cases} \tag{1}$$

With

$$\delta_i = \begin{cases} 1 & \text{event realized,} \\ 0 & \text{censored subject.} \end{cases}$$

- Case of exo aequo :

In the case of application, we are confronted with the presence of events of different natures, we consider that the uncensored observations take place before the censored ones, we have:

$$\hat{S}(t) = \begin{cases} \Pi_{T_{(i)} \leq t} \left(1 - \frac{d_i}{n_i}\right) = \Pi_{t_i \leq t}(1 - q_i) = \Pi_{t_i \leq t} p_i & \text{si } t \geq t_1 \\ 1 & \text{si } t < t_1 \end{cases} \tag{2}$$

The Kaplan-Meier estimator is convergent formed a piecewise constant function, right continuous and left limit, with a jump at each observed time of death.

1.2. The estimator of Nelson-Aalen[1]

Estimation of the cumulative hazard rate $\wedge(t)$ through the Nelson and Aalen method is based on the following equation :

$$\wedge(u + du) - \wedge(u) \approx \lambda(u)du \tag{3}$$

Equation (3) is asymptotic to the probability of dying within a small time interval after time u, it is given by:

$$P(u < X \leq u + du / X > u) = \frac{P(u < X \leq u + du)}{P(X > u)} \tag{4}$$

It is natural to estimate the ratio (4) by:

---

[1] Cet estimateur à été proposé par Nelson dans l'année 1972 et ensuite par Aalen en 1978.

$$\frac{[N(u+du)-N(u)]}{Y(u)} \tag{5}$$

Thus, the probability (5) is approximated by the frequency:

$$\frac{[N(u+du)-N(u)]}{Y(u)} = \frac{dN(u)}{Y(u)} \tag{6}$$

Where $dN(u)$ is the number of "events" observed in the interval $]u, u+du]$ and $Y(u)$ is the number of "individuals" at risk at time u.

By summing the quantities (6) over all the subintervals forming $(0, t]$ and making these intervals tend to 0, so that each contains only one event. Thus, the Nelson-Aalen estimator is defined by:

$$\widehat{\Lambda}_{NA}(t) = \int_0^t \frac{dN(u)}{Y(u)}$$

which can also be written, if we note the moments $t_1, t_2, \ldots, t_N$ orders of occurrence of the event for the N individuals (jump process):

$$\int_0^t \frac{dN(u)}{Y(u)} = \sum_{\{i, t_i \le t\}} \int_{t_{i-1}}^{t_i} \frac{dN(u)}{Y(u)} = \sum_{\{i, t_i \le t\}} \frac{N(t_i) - N(t_{i-1})}{Y(t_i)}$$

$$= \sum_{\{i, t_i \le t\}} \frac{\Delta N(t_i)}{Y(t_i)} = \widehat{\Lambda}_{NA}(t)$$

Or on $]t_{i-1}, t_i]$ the process $Y(u)$ is constant equal to $Y(t_i)$ and $\Delta N(t_i)$ is the number of events observed in the interval $]t_{i-1}; t_i]$ i.e. in t_i :

$$\Delta N(t_i) = N(t_i) - N(t_{i-1})$$

According to the last equation, it is possible to find an estimator for the instantaneous hazard rate $\lambda(t_i)$ in t_i , such as the estimator $\widehat{\Lambda}_{NA}(t)$ of Nelson-Aalen, is written as follows:

$$\widehat{\Lambda}_{NA}(t) = \sum_{\{i, t_i \le t\}} \hat{\lambda}(t_i)$$

We put:

$$\hat{\lambda}(t_i) = \frac{d_i}{n_i}$$

We find the final form of this estimator as follows:

$$\widehat{\Lambda}_{NA}(t) = \sum_{\{i, t_i \le t\}} \hat{\lambda}(t_i) = \sum_{\{i, t_i \le t\}} \frac{d_i}{n_i} \tag{7}$$

Proposition 1. An estimator of the variance of $\widehat{\Lambda}_{NA}(t)$ is :

$$var[\widehat{\widehat{\Lambda}_{NA}}(t)] = \sum_{i:t_i \le t} \frac{d_i}{n_i^2} \tag{8}$$

1.3. Harrington and Fleming estimator of survival

From the relationship $S(t) = exp(-\wedge(t))$ and from the Nelson-Aalen estimator, we can deduce another estimator of the survival function:

$$\hat{S}_{HF}(t) = exp\left(-\hat{\wedge}(t)\right) = exp\left(-\sum_{\{i,t_i \leq t\}} \frac{d_i}{n_i}\right)$$

$$= \prod_{i;T_i \leq t}^{n} e^{-\frac{d_i}{n_i}}$$

$$\approx \prod_{i;T_i \leq t}^{n} \left(1 - \frac{d_i}{n_i}\right), si \frac{d_i}{n_i} \to 0 \qquad (9)$$

where d_i and n_i are the number of deaths and individuals at risk in T_i. By applying a limited expansion, we find the Kaplan-Meier estimator. Using the delta-method $(Var(f(Z))) \sim [f'(E(Z))]^2 Var(Z)$, one can obtain an estimator of the variance of this estimator :

$$\widehat{Var}\left(\hat{S}_{HF}(t)\right) = \left(\hat{S}_{HF}(t)\right)^2 \widehat{Var}\left(\hat{\wedge}(t)\right)$$

$$= exp\left(-2\sum_{i;T_i \leq t}^{n} \frac{d_i}{n_i}\right) \times \left(\sum_{i;T_i \leq t}^{n} \frac{d_i}{n_i^2}\right) \qquad (10)$$

2. the Bayesian approach and classical non-parametric methods

2.1. On the Bayesian estimation of the Kaplan-Meier model

The calculation of the Kaplan-Meier estimator requires individual data with exact dates of the events but this information is not always accessible, to deal with this problem it is possible to use the link with the parametric approach. Indeed we assume that the number of deaths in the time interval is a realization of a Binomial law written by:

$$d_i \sim \beta in(n_i, q_i)$$

Exits in the intervals $[t_i, t_{i+1}[$ being independent of each other, we therefore find that the likelihood of this model is written:

$$f(d_i/q_i) = \prod_{i=1}^{m} C_{n_i}^{d_i} q_i^{d_i}(1-q_i)^{n_i-d_i}$$

the log-likelihood is therefore written:

$$lnL = \sum_{i=1}^{m}\left[C_{n_i}^{d_i} + d_i ln(q_i) + (n_i - d_i)ln(1-q_i)\right]$$

first-order conditions $\frac{\partial}{\partial} lnL = 0$ then lead to the estimators :

$$\hat{q}_i = \frac{d_i}{n_i} ,$$

this equation is the idea of the Kaplan-Meier method, where the probability of surviving until the instant t_i is the probability of survival t_i knowing that we were alive in t_{i-1}.

From a Bayesian perspective, we assume an a priori for q_i, as the natural conjugate of a Binomial law is a beta law, we set:

$$q_i \sim \mathcal{B}e(\alpha, \beta),$$

2.2. On Bayesian estimation of the Nelson-Aalen model

The estimator $\widehat{\Lambda}_{NA}(t)$ of Nelson- Aalen (1972,1978), is written as follows:

$$\widehat{\Lambda}_{NA}(t) = \sum_{\{i, t_i \leq t\}} \hat{\lambda}(t_i)$$

We put:

$$\hat{\lambda}(t_i) = q_i$$

Such that the discretized version is given by a prior

$$q_i \sim \mathcal{B}e(\alpha, \beta),$$

We find the final form of this estimator as follows:

$$\widehat{\Lambda}_{NA}(t) = \sum_{\{i, t_i \leq t\}} \hat{\lambda}(t_i) = \sum_{\{i, t_i \leq t\}} q_i \qquad (11)$$

2.3. On the Bayesian estimation of the Harrington and Fleming model

The proposed model is written as follows

$$q_i = \frac{exp(\mu_i)}{1 + exp(\mu_i)}$$

$$\mu_i \sim \mathcal{N}(\vartheta; \tau)$$

$$\vartheta \sim \mathcal{N}(0; 0{,}001), \tau \sim HC(B)$$

$$B \sim Uniforme(0; \mathcal{T}), \text{on pose}: \mathcal{T} = 100$$

$$\hat{S}_{HF}(t) = exp(-\widehat{\Lambda}(t)) = \prod_{i; T_i \leq t}^{n} e^{-\frac{d_i}{N_i}}$$

3. The Bayesian deviance criterion and the choice of model

Spiegelhalter et al (2002) propose a generalization of AIC and BIC (Bayesian Information Criterion) because asymptotic justification is not appropriate in hierarchical models. The generalization is based on the posterior distribution of the deviance statistic,

$$D(\theta) = -2(logf(x/\theta) - logh(x)) \qquad (12)$$

$f(x/\theta)$ is the likelihood function, $h(x)$ is a data normalization function. These authors suggest summarizing the fit of a model by the posterior expectation of the deviance (residual deviance), $\bar{D} = E_{\theta/x}[D]$, and model complexity p_D called the effective number of parameters, i.e. the number of parameters making up the model. In the case of Gaussian models, it can be shown that a reasonable definition of p_D is the expected deviation minus the deviation calculated from the posterior estimates of the parameters,

$$p_D = E_{\theta/x}[D] - D(E_{\theta/x}[\theta]) = \bar{D} - D(\bar{\theta})$$

The information deviation criterion (DIC) is defined as follows:

$$DIC = \bar{D} + p_D = \bar{D} + \bar{D} - D(\bar{\theta}) = 2\bar{D} - D(\bar{\theta}) \qquad (13)$$

This criterion is more satisfactory than the previous ones because it takes into account the a priori information and integrates a natural penalization factor into the log-likelihood (Robert (2006)). The model retained is the one with the smallest DIC value. Spiegelhalter et al (2002), propose that a difference of 5 or 10 in the value of DIC is generally negligible. Regarding the variance of DIC by the Monte Carlo method, one calculates the DIC a few times and using different seeds of random numbers to get a rough idea of the variability in the estimates. With a large number of independent repeat DICs $\{DIC_l, l = 1, ..., N\}$, $Var(DIC)$ is estimated by:

$$\widehat{var}(DIC) = \frac{1}{N-1} \sum_{l=1}^{N} (DIC_l - \overline{DIC})^2 \qquad (14)$$

4. Application of duration models in econometrics

- **Presentation of data**

The data we have relate to a filtered sample of 1,064 unemployed people registered with the local employment agency of Ain El Benian, over the period from January 1, 2011 to July 15, 2013. Distinguishing those who found a employment, the placement of the unemployed during this period gives rise to 875 right-censored observations. In this case, the variable i represents the indication that the ith unemployed person has accessed a job after his daily period of unemployment t_i.

Figure (1): Kaplan-Meier survival functions for aggregate unemployment duration.
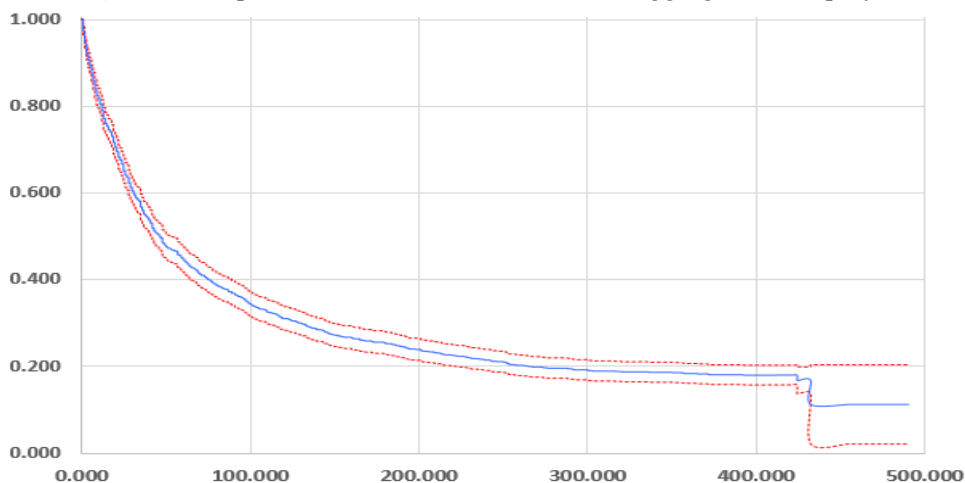
Source: Developed by us, using Excel program.

From figure (1), we notice that at the beginning of the curve, 100% of the individuals in the sample are unemployed. After approximately 2 months of registration with this agency, 50% of individuals were placed in the labor market. However, the exit from

unemployment for the rest of the individuals in the sample is spread over a long period, for some it even exceeds one year. In general, according to the unemployment duration curve, we deduce that the probability of leaving unemployment for those registered at the Local Employment Agency of Ain el Benian becomes very low for an unemployed person who exceeds more than a year of unemployment.

Figure (2) : Bayesian survival functions according to the Kaplan Meier method with a prior beta (0.01,0.01) and the Fleming-Harrington method.
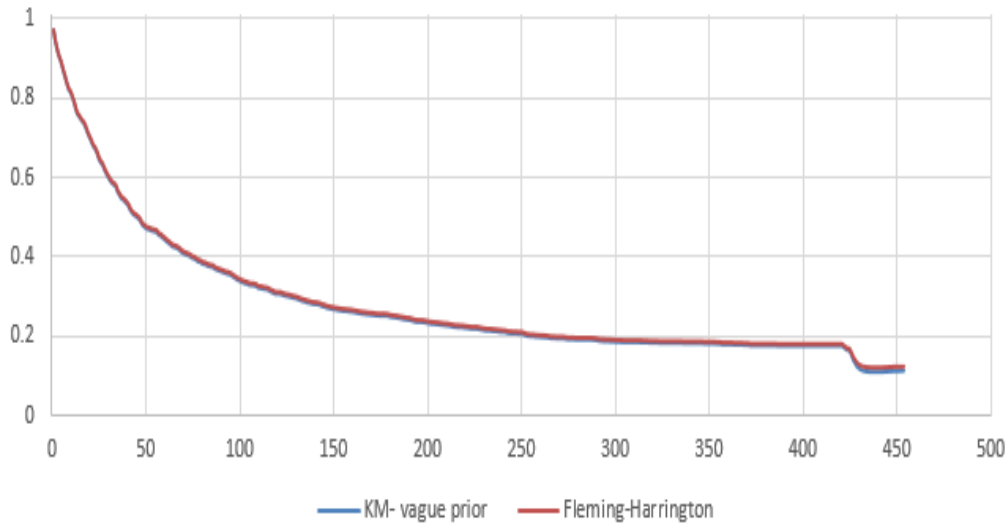
Source: Developed by us, using Excel program.

This figure (2) informs us that the median duration of exit from unemployment is 45 days in the two methods. Thus, the two estimation methods) are approximately identical despite the number of individuals being large, and the difference that exists is in the final durations as Figure (4) shows.

Figure (3) : Bayesian survival functions for durations longer than 257 days and according to the Kaplan Meier method with a prior beta(0.01,0.01) and the Fleming-Harrington method.
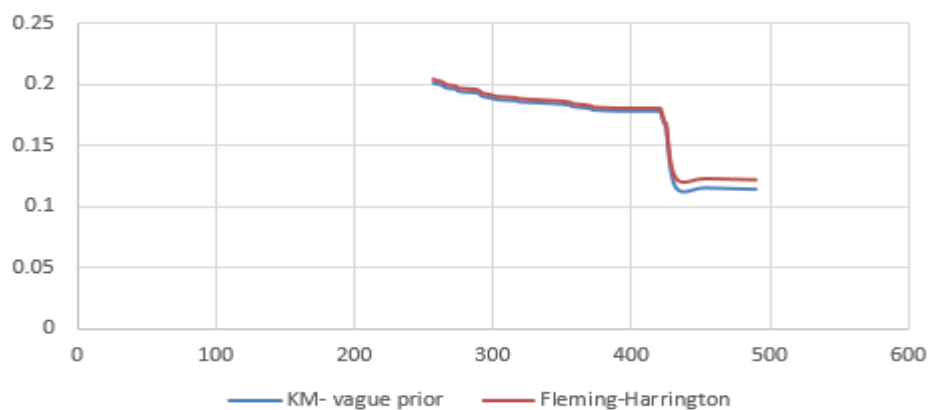
La source: Developed by us, using Excel program.

In figure (3), there is a slight difference between the survival curves according to the different methods used. Consequently, we use the DIC (the information deviance criterion) to choose the best model.

Table (1): Comparison between the Bayesian survival functions according to the Kaplan Meier method with a prior beta (0.01, 0.01) and the Fleming-Harrington method.

	Kaplan Meier's method	The Fleming-Harrington method
DIC	1157	1140

Source: Developed by us, using Excel program.

In the accuracy aspect we find that the Fleming-Harrington method presents a higher efficiency, because the difference in the information deviation criterion is greater than 5.

5. Results

The relevance and effectiveness of the Bayesian approach as a guide to scientific reasoning in the face of uncertainty have long been recognized, in many situations of random experiments, the practitioner has information on the phenomenon studied, opinions of researchers and experts, professional experiences, and acquired observations, these two methods explained in our article allow the use of this information and in different ways.

6. Discussions and conclusion

The objective sought in our contribution is the improvement of the inferential phase in the estimation of nonparametric survival times and in the presence of censoring. This is possible according to a link With the parametric approach, we speak of a discretized version of the classical nonparametric estimators, this passage makes it possible to overcome the problem of force of habit when using complex Bayesian methods such as those nonparametric methods requiring a high mathematical and theoretical package. We have shown this efficiency with the application of Gibbs sampling. The use of the MCMC method is relatively easy to implement, it provides a set of techniques very suitable for estimating complex models with several parameters or with a hierarchical structure.

Referrals and references:

– CARLIN, B.P., GELFAND, A.E et SMITH, A.F.M. (1992). *Hierarchical Bayesian Analysis of Change point Problems*. Appl. Statist. 41(2), 389-405.

– CLAYTON, D.G. (1978). *A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence*. Biometrika, 65, 141 – 151.

– FERGUSON, T.S. (1973). *A Bayesian analysis of some nonparametric problems*, The Annals of Statistics. 1(2), 209-230.

– Fleming, T. R et Harrington, D. P. (1984). *Nonparametric estimation of the survival distribution in censored data*. *Communications in Statistics-Theory and Methods*. 13(20), 2469-2486.

– FLORENS, J.P et ROLIN, J.M. (2001). Simulation of posterior distributions in nonparametric censored analysis. International Statistical Review. 67(2), 187-210.

– Geman, S et German, D. (1984). *Stochastic relaxation, Gibbs distribution and the Bayesian restoration of image*. IEEE Trans. Pattern Anal. Mach. Intell. 6,721-741.

– Haldane, J. (1931). *A note on inverse probability*. Proc. Cambridge Philos .

– Kalbfleisch, J.D et Prentice, R.I. (1980). *The Statistical analysis of failure time data*. John Wiley & Sons. New York.

– KALBFLEISCH, J.D. (1978). *Nonparametric bayesian analysis of survival time data*. Journal of the Royal Statistical Society, Series B. 40(2), 214-221.

– Kaplan, E, L et Meier, P .(1958). *Nonparametric estimation from incomplete observations*.J.Amer. Statist. Assoc. 53, 457–481. MR0093867 (20:387).

– Khizanov, V. G et Maïboroda, R .(2015). *A modified Kaplan-Meier estimator for a model of mixtures with varying concentration*s. Theor. Probability and Math. Statist. 92 (2016 Field, C. A.

– Novick, M. et Hall, W .(1965). *A Bayesian indifference procedure*.J. American.

– Robert, C.P .(2006). *Le choix Bayésien : principes et pratiques*. Springer.

– Ronchetti, E et Ronchetti, E. M .(1990). *Small sample asymptotics*. Ims.

– Rossa, A et Zieliński, R .(2006). *A simple improvement of the kaplan-meier estimator*. Communications in statistics - theory and methods. 31(1), 147-158, doi: 10.1081/sta-120002440.

– Shafiq, Mohammad., Shah, Shuhrat et Alamgir, M .(2007). *Modified Weighted Kaplan-Meier Estimator*. Pakistan Journal of Statistics and Operation Research, 3,39-44.

– Spiegelhalter, D. J., Best, N., Carlin, B.P et Van der Linde, A .(2002). *Bayesian measures of model complexity and fit*. Journal of the Royal Statistical Society, Series B, 64, 583–640.Statist. Assoc., 60, 1104–1117.

The importance of selecting Bayesian nonparametric survival models: application to the estimation of unemployment durations

HAMIMES Ahmed

Appendices

A1. The Fleming-Harrington Bayesian model:

```
model
{
for (i in 1:m) {
d[i]~dbin(q[i],n[i])
q[i]~dbeta(0.01,0.01)
}
for (i in 1:m){
ce[i]~dbin(qc[i],n[i])
}
for (i in 1:m){
qc[i]~dbeta(0.01,0.01)
}
for (i in 1:m){
pp[i]<-1-q[i]
}
n[1]<- 1064
for(i in 2:m){
n[i]<-n[i-1]-d[i-1]-ce[i-1]
}
for (i in 2:m){
s[i]<- exp(-sum(q[1:i]))
}
s[1]<-pp[1]
}
list( m=227,
ce=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0,36,39,29,42,30,6,5,0,1,1),
d=c(28,31,23,18,14,20,18,18,16,8,14,15,20,12,7,9,5,11,15,12,11,13,7,10,16,11,6,13,11,9,9,7,5,4,
14,9,9,4,5,6,7,12,7,5,3,5,5,11,5,5,3,1,3,1,2,1,8,2,5,4,5,4,4,5,3,1,3,3,6,3,3,4,3,4,1,4,4,1,4,2,1,3,2,1
,1,5,4,4,2,2,1,2,3,4,3,4,2,3,2,2,1,3,1,1,2,4,1,1,1,1,1,3,3,2,3,1,2,2,1,2,2,1,1,2,4,2,1,2,1,3,1,2,3,1,2,
2,1,1,1,2,2,1,2,1,1,3,1,1,2,1,1,2,1,2,2,1,2,1,2,1,2,1,3,1,1,1,2,1,2,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,1
,1,1,2,2,1,1,1,1,1,1,1,1,1,1,1,1,1,1,2,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,1,0,1,0,0)),
list(q=c(0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.
5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0
.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,
0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5
,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.
5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0
.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,
0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5
,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5)),
list(q=c( 0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
```

```
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9))
```

A2. Kaplan Meier's Bayesian model:

```
model
{
for (i in 1:m) {
d[i]~dbin(q[i],n[i])
q[i]~dbeta(0.01,0.01)
}
for (i in 1:m){
ce[i]~dbin(qc[i],n[i])
}
for (i in 1:m){
qc[i]~dbeta(0.01,0.01)
}
for (i in 1:m){
p[i]<-1-q[i]
}
n[1]<- 1064
for(i in 2:m){
n[i]<-n[i-1]-d[i-1]-ce[i-1]
}
for (i in 2:m){
s[i]<- s[i]*p[i]
}
s[1]<-p[1]
}
list( m=227,
ce=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
,0,0,0,0,0,0,0,0,0,0,0,0,36,39,29,42,30,6,5,0,1,1),
d=c(28,31,23,18,14,20,18,18,16,8,14,15,20,12,7,9,5,11,15,12,11,13,7,10,16,11,6,13,11,9,9,7,5,4,
14,9,9,4,5,6,7,12,7,5,3,5,5,11,5,5,3,1,3,1,2,1,8,2,5,4,5,4,4,5,3,1,3,3,6,3,3,4,3,4,1,4,4,1,4,2,1,3,2,1
,1,5,4,4,2,2,1,2,3,4,3,4,2,3,2,2,1,3,1,1,2,4,1,1,1,1,1,3,3,2,3,1,2,2,1,2,2,1,1,2,4,2,1,2,1,3,1,2,3,1,2,
2,1,1,1,2,2,1,2,1,1,3,1,1,2,1,1,2,1,2,2,1,2,1,2,1,2,1,3,1,1,1,2,1,2,1,1,1,1,1,1,1,1,1,1,1,1,1,2,1
,1,1,2,2,1,1,1,1,1,1,1,1,1,1,1,1,1,2,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,1,0,1,0,0))
list(q=c(0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.
5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0
```

.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,
0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5
,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.
5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0
.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,
0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5
,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5))
list(q=c(0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,
0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9,0.9))