Journal of Economic & Financial Research
ISSN : 2352– 9822
E-ISSN : 2588-1574
Volume 7 / Issue 2 / December 2020
OEB Univ. Publish. Co.

Bayesian survival study on the integration impact of the unemployed at the Ain El Benian local employment agency (January 2011-July 2013)

دراسة مدة البقاء وفق النهج البايزي حول تأثير اندماج العاطلين عن العمل في وكالة التوظيف المحلية في عين البنيان

(**يناير** 2011 -**يوليو** 2013)

Ahmed Hamimes*

ahmedhamimes@yahoo.com

PHD Student, National Higher School of statistics and applied economics, Tipaza, Algeria

Benamirouche Rachid

rbena2002@hotmail.com

Professor, National Higher School of statistics and applied economics, Tipaza, Algeria

Received: 13/10/2020 Accepted: 15/11/2020 Published: 31 /12/2020

Abstract:

In this article, the Kaplan Meier model is used in the survival analysis and according to a Bayesian conception in the context of using the posterior mean approach to determine the mean value of the risk in the study of the impact of the unemployed registered with the local employment agency of Ain El Benian (January 2011-July 2013). This method makes it possible to probabilize the average value of the risk of chance and to compare it with the national unemployment rate; in this way, it is possible to measure the integration value of the unemployed at this agency.

Keywords: *The posterior mean approach, Kaplan Meier model, Unemployment.*

Jel Classification Codes: *C11, C41, E24.*

***Corresponding author:** Ahmed Hamimes

الملخص:

في هذه المقالة، يتم استخدام نموذج كابلان ماير في تحليل البقاء ووفقًا لمفهوم بايز في سياق استخدام نهج المتوسط اللاحق لتحديد القيمة المتوسطة للمخاطر في دراسة تأثير العاطلين عن العمل المسجلين لدى وكالة توظيف عين البنيان (يناير 2011 – يوليو 2013). تتيح هذه الطريقة إمكانية توقع القيمة المتوسطة لخطر البطالة ومقارنتها بمعدل البطالة الوطني؛ وبهذه الطريقة يمكن قياس قيمة اندماج العاطلين عن العمل في هذه الوكالة.

كلمات مفتاحية: منهج المتوسط اللاحق، نموذج كابلان ماير ، البطالة.

تصنيف *Jel*: *E24, C41, C11*.

I. Introduction

Bayesianism has been the fundamental schism in statistics, torn between Bayesian and Classical. Bayesian statistics can be presented as a generalization of the frequentist approach. The Bayesian statistician considers the parameter of the statistical model $f\ (x\ /\ \theta)$ to be uncertain. He will therefore seek to quantify his uncertainty by modeling all the available information. This is what makes all the difference since it considers θ a random parameter and it consists in probabilizing the space of the parameters Θ, that is to say in taking into account subjective ideas (a priori: opinion of analysts, experts ...) on the parameters.

The choice to probabilize the space of states is truly revolutionary. Statisticians draw an hermetic border between the notion of unknown parameter and random parameter. They defend this point of view on the basis that, although in some settings, the parameter is obtained under the simultaneous action of several factors and can thus appear as (partially) random, for example in models of mixtures or in physics quantum, in most cases it cannot be seen as the result of a random experiment.

The following table summarizes the differences in conceptions of probability between the paradigms.

Table.1: conceptions of probability concerning controversial points.

Conception	Problèmes et points controversés
Subjective conception.	• The estimate of the probability can vary between different persons. • The probability does not necessarily have to correspond to reality. • This is an individual probability that we cannot necessarily verify empirically. • The attribution of a probability can be very difficult due to lack of information.
Classic conception.	• Equiprobability is a constraint very strict. • Therefore the design cannot be still applied in the real world. • Laplace's design cannot be applied when there is not afinite universe (Ω). It's also easier if the universeis small. • Before properly defining a notion of probability, we already presuppose it.

Source: Developed by us.

Survival data is the time between the start of an experience and the onset of an event. The simplest event is death; however, other events, such as the onset of disease or an epidemic, are covered by the given term of survival. In manufacturing, it could be the breakdown of a computer or the time it takes for a person to get a job in the economy. The event is in many situations the passage from one state to another. To begin with, death is the transition from the "living" state to the "dead" state. The onset of the disease is the transition from "healthy" to "sick".

The first field of application of the analysis of duration data was in the biomedical sciences where it is used either for a therapeutic trial or for epidemiological studies. Then, this data has been widely used in the social sciences for the purpose of analyzing the weather for events such as job changes, marriage, childbirth, etc. These data are also a tool used in many fields, namely insurance (the length of time between two claims of the same nature), reliability (the time between successive breakdowns), demography (the length of human life), economics (the duration of the work stoppage, the duration of the strike, the lifespan of a company), finance (the analysis of credit risks, the duration between two transactions on the same security).

In the case of semi-parametric models, we find a kind of compromise between the two approaches of parametric and non-parametric survival. The real probability law of the observations is assumed to belong to a class of laws that are partly dependent on parameters, and partly written as a non-parametric function. This approach is widely used in survival analysis, notably through the regression model of Cox (1972) in the frequentist approach, several authors have discussed the use of Bayesian inference in the analysis of semi-survival. parametric (Ferguson 1973; Kalbfleisch 1978; Kalbfleisch and Prentice 1980; Clayton 1978). The parametric approach stipulates that the real probability law of observations belongs to a particular class of laws, which depend on a certain (finite) number of parameters. The advantage of this approach is the expected facilitation of the parameter estimation phase, as well as the obtaining of confidence intervals and the construction of tests. The disadvantage of the parametric method is the mismatch that may exist between the phenomenon studied and the model selected. Parametric models have taken a significant place in the study of Bayesian survival, we give here some references in the medical or public health field (Grieve, 1987; Achcar et al., 1987; Achcar et al., 1985; Chen et al., 1985; Dellaportas and Smith, 1993; Kim and Ibrahim, 2001), the book by Ibrahim et al. (2001) provides a clear description of Bayesian survival models in general, and parametric models in particular. We can also cite the Bayesian method for breaking point models introduced by Carlin, Gelfand and Smith (1992) in parametric modeling. The non-parametric approach does not require any assumptions about the actual probability law of the observations, and this is its main advantage. It is therefore a problem of functional estimation, it implies, for example, the survival function, which is continuous, will be estimated by a discontinuous function. The disadvantage of such an approach is the need to have a large number of observations, the problem of estimating a functional parameter being delicate since it belongs to a space of infinite dimension. Florens and Rolin (2001) mainly demonstrated in nonparametric modeling that the simulation estimates of the Dirichlet process and nonparametric Bayesian inference give good results compared to classical methods.

In the classical approach, there are two very well-known non-parametric methods for estimating $S(t)$ in the presence of censorship, i.e. the actuarial method and the Kaplan-Meier method (also called the product-limit), the difference is that Kaplan-Meier method allows to estimate the survival functions, without requiring regular time intervals. In this contribution, we will give a Bayesian alternative to the classical Kaplan Meier estimator based on the beta distribution. We find this important to give a Bayesian estimator based on these two classical conceptions for several reasons, firstly and in medical science for example, we very often have to deal with small samples. There are several reasons for this. The most common is the rarity of the disease or the difficulty in bringing together patients with the same biochemical parameters. In addition, we have very often censored data. Therefore, a small sample size does not allow us to use classical statistical methods or when they are used they can give us too general and even false results. The second reason by using the Bayesian method, we get results for both censored and uncensored cases. This is why our survival curve is smoother and we don't have such rapid jumps for the probability of survival. In addition, for some observations, the interval is even smaller. The resort to the Bayesian method in the case of the Kaplan Meier estimator simplifies the use of this approach over other methods (the methods based on randomized measurement) and addresses the problem of usual strength in the use of frequentist estimates. This method makes it possible to probabilize the average value of the chance risk and compare it with the national unemployment rate; in this way, it is possible to measure the integration value of the unemployed at this agency.

II. The Bayésien paradigm[1]

The concept of "Bayesian statistics" starts from the neologism "Bayesian", taken from the name of Thomas Bayes, who introduced the theorem which now bears his name in a posthumous article of 1763, 250 years ago. This theorem expresses a conditional probability in terms of the inverse conditional probabilities, if A and E are events such that $P(E) \neq 0$, $P\,(A/E)$ and $P(E/A)$ are related by:

$$P(A/E) = \frac{P(E/A)P(A)}{P(E/A)P(A) + P(E/\bar{A})P(\bar{A})} = \frac{P(E/A)P(A)}{P(E)} \qquad (1)$$

In particular, if an event E, can be the result of a cause A or a cause B, the Laplace principle can be written:

$$\frac{P(A/E)}{P(B/E)} = \frac{P(E/A)}{P(E/B)}$$

when $P\,(B) = P\,(A)$. Obtaining this result from the axioms of Probability Theory is trivial. It is, however, the most important conceptual step in the history of statistics, constituting the first inversion of probabilities. This equation expresses the fundamental fact that, for two equiprobable causes, the ratio of the probabilities for a given effect is equal to the ratio of the probabilities of these two causes. Also, this theorem is also a discounting principle, since it describes the updating of the likelihood of A from $P\,(A)$ to $P\,(A \mid E)$, once E has been observed (Robert, 2006).

[1] The word paradigm, which is a grammatical term, is misused here as a synonym for "thought model" or "principles".

Bayes[2] used the continuous version of this theorem, which states that, for:

- $f(x/\theta)$ the density function of x knowing the value of the random variable $\theta,$
- $\pi(\theta)$ models the density function a priori on $\theta,$
  the posterior distribution of θ conditional on x is:

$$\pi(\theta/x) = \frac{f(x/\theta) \times \pi(\theta)}{\int_{\Theta} f(x/\theta) \times \pi(\theta) d\theta} = \frac{f(x/\theta) \times \pi(\theta)}{m(x)} \qquad (2)$$

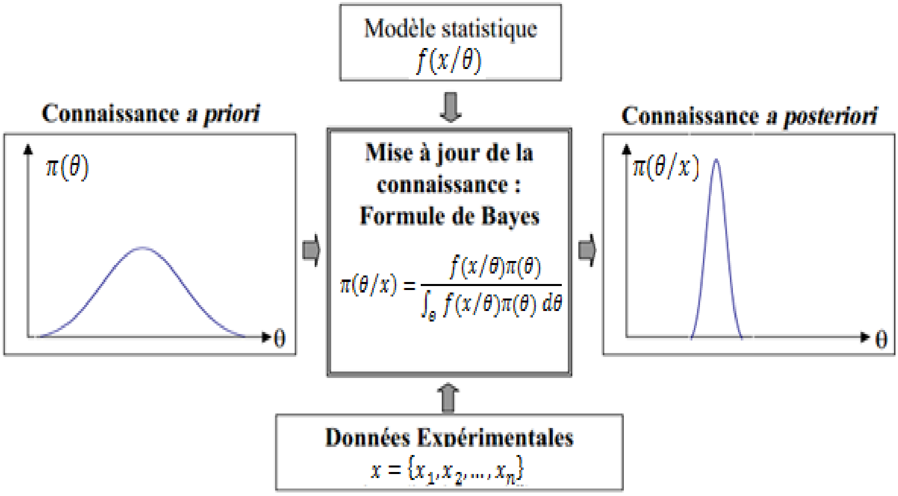
Once we have the data, the quantity $m\ (x)$ is a normalization constant which guarantees that
$\pi\ (\theta\ /\ x)$ is indeed a probability distribution.
We can write:

$$\pi(\theta/x) \propto f(x/\theta) \times \pi(\theta) \qquad (3)$$

Définition 1. (The Bayesian statistical model).
We call Bayesian statistical model, the data of a parameterized statistical model $(\mathfrak{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$ with $f(x/\theta)$ density of P_θ and a law $\pi\ (\theta)$ on the parameter.
The fundamental concept of the Bayesian paradigm is a posteriori distribution. It gives the degree of credibility of the range of possible values of the state of nature knowing that the experimental result x has occurred, and therefore automatically sets the inversion of the probabilities, while including the requirements of the likelihood principle.

[2] Thomas Bayes 'work published in 1763 after his death because he was not sure of his conclusions, some authors interpret the fact that Bayes' formulas were published posthumously (in 1763) out of fear of sacrilege : Thomas Bayes was indeed a clergyman and the application of his formula to the search for the ultimate causes of an event could have led to the probability of the existence of God ...

Figure.1: The Bayesian paradigm: summary (source: Boreux et al (2000)).



Source: Developed by us .

According to figure 1, the a posteriori distribution represents the updating of the information available on the parameter θ in view of the information contained in the likelihood $f(x/\theta)$, while $\pi\,(\theta)$ represents the knowledge available a priori. The Bayesian framework is therefore presented as a formalized theory of learning by experience.

III. The Bayesian conception of the Kaplan Meier estimator

The Kaplan-Meier estimator (1958) is a functional method for estimating the survival function is written

$$\hat{S}(t) = \begin{cases} \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right) = \prod_{t_i \le t} (1 - q_i) = \prod_{t_i \le t} p_i & \text{si } t \ge t_1 \\ 1 & \text{si } t < t_1 \end{cases} \quad (4)$$

$t_i$represents the follow−up time since inclusion in the study for each individual i.

$d_i$is the number of deaths at time t_i.

$n_i$is the number of subjects at risk of presenting the studied event at timet_i, i.e. the number of patients who have not yet undergone the event nor the censorship just before t_i.

In a Bayesian view it is assumed that the number of deaths in the interval of time is a Binomial distribution given by

$$d_i \sim \beta in(n_i, q_i) \quad (5)$$

In this discretized approach of the Kaplan Meier model, the risk of death estimated by the likelihood method is:

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

The parameters $q_{i,}$ in the Bayesian framework are random variables, and when the distribution used in the case of the proportions is that of Beta, we set:

$$q_{i\sim} beta(\alpha, \beta) \quad (6)$$

The prior law is considered to be weak informative, it provides solutions in the use of algorithms. We pose:

$$q_i \sim \beta(0.01,0,01) \quad (7)$$

In this article we use the hierarchical method in the construction of a priori distributions, then:

$$q_i \sim \beta(\alpha, \beta) \quad (8)$$

and

$$\alpha \sim Gamma(0.0001,0,0001) \quad (9)$$

$$\beta \sim Gammma(0.0001,0,0001) \quad (10)$$

This representation is based on the hierarchical method in the selection of a priori laws; such as the hierarchical bayesian approach tends to model the lack of information on the parameters of the prior distribution by a new distribution on these parameters. In other words, these hyper parameters are distributed using hyper a priori laws, and must therefore be estimated without resorting to the data, in order to calculate the average a posteriori risk.

IV. the average value of the risk in the case of data of weakly and strongly censored durations

The Kaplan Meier estimator is a generalization of the notion of empirical distribution function, is based on the following idea: to survive after time t is to be alive just before t and not to die at time t.

We use the discretized version of this estimator such that the number of deaths in the time interval is a realization of a Binomial distribution is written by:

$$d_i \sim \beta in(n_i, q_i), \tag{11}$$

When the outputs in the intervals $[t_i, t_{i+1}[$ being independent of each other, we write:

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

through the logarithm function we find:

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

using the maximum likelihood method, the risk of death estimated by:

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

A Bayesian vision linked to the estimation of the a posteriori mean value of the risk, we use the hierarchical method in the construction of the a priori laws, then:

$$q_i \sim \beta(\alpha, \beta) \tag{12}$$

and

$$\alpha \sim Gamma(0.0001{,}0{,}0001) \tag{13}$$

and

$$\beta \sim Gammma(0.0001{,}0{,}0001) \tag{14}$$

in this model we assume that $q_{i,}$ is constant and follows a saddle distribution

$$q_{moyenne\sim} beta(\alpha, \beta) \tag{15}$$

To check the quality of the Kaplan Meier Bayesian model in this operation we use the observed values of d_i to form the statistic

$$\chi^2_{obs} = \sum_i \frac{(d_{obs.i} - \mu_i)^2}{\sigma_i^2} = \sum_i \frac{(d_{obs.i} - (q_i * n_i))^2}{(q_i * n_i) * (1 - (q_i)} \tag{16}$$

We then generate predicted values of d_i from its posterior predictive distribution, and construct an analogous statistic:

$$\chi^2_{rep} = \sum_i \frac{(d_{rep.i} - \mu_i)^2}{\sigma_i^2} = \sum_i \frac{(d_{resp.i} - (q_i * n_i))^2}{(q_i * n_i) * (1 - (q_i)} \tag{17}$$

evaluating the posterior distributions of $D(d_{obs.i}, q_i)$ and $D(d_{rep.i}, q_i)$ provides individual and aggregate measures of goodness of fit that can be described graphically or in using tail region probabilities called a posteriori predictive values (Meng, 1994).

$$p - value \equiv P\left[D(d_{rep.i}, q_i) \ \geq D(d_{obs.i}, q_i)/d_i\right]$$

Gelman, Meng and Stern (1996) recommend calculating the "predictive p-value"

$$p - value \equiv P\left[D(d_{rep.i}, q_i) \geq D(d_{obs.i}, q_i)/d_i\right]$$

$$= \int\int I_{[D(d_{rep.i}, q_{constant}) \geq D(d_{obs.i}, q_{constant})]} f({d_i}^{rep}/q)\, \pi(q_i/d_i) {d_i}^{rep} d\theta \tag{18}$$

where I is the indicator function.

This integral can be approximated by sampling ${q_i}^k$ from the posterior distribution of q_i, the same thing for ${d_i}^{rep}$from the distribution $f({d_i}^{rep}/q_i)$. In the result we find

$$p - value = \sum_{k=1}^{T} I\left[D\left({d_i}^k, {q_i}^k\right) \geq D\left(d_i, {q_i}^k\right)\right]/T \tag{19}$$

In the case of data of heavily censored durations, the number of deaths in the time interval is a Binomial distribution given by:

$$d_i \sim \beta in(n_i, q_i) \tag{20}$$

and

$$q_{i\sim} beta(\alpha, \beta) \tag{21}$$

In the case of data of strongly censored durations, it is not necessary to use the hierarchical version of the a priori law therefore: $\alpha = \beta = 0.01$.

In this model we assume that$q_{i,}$ is constant and follows a saddle distribution

$$\widehat{q}_{moyenne} = \hat{q}_{j,r} \quad (22)$$

we assume the categorial density:

$$r \sim cat(w_i); \sum_{i=1}^{n} w_i = 1$$

we also assume weights for the parameters $w_1, w_2, \dots, w_n$. Such as:

$$w_1 + w_2 + \cdots + w_n = 1$$

the parameters $w_1, w_2, \dots, w_n$represent a function of the censored data having lower weights than those of the uncensored data, then the average risk value is realized via a point approximate approach of the uncensored risks.

V. Application

1. unemployment durations and the model used

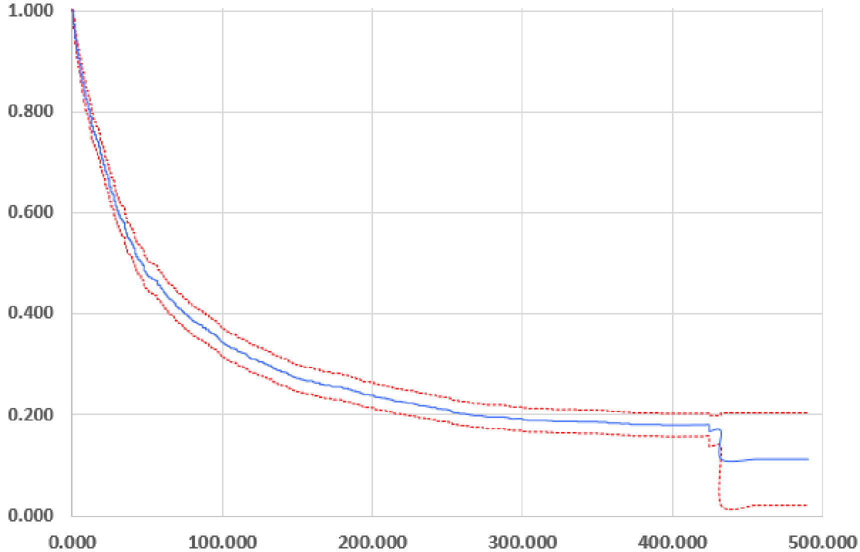
One of the most effective tools for analyzing lifespans is certainly the survival or stay function estimator. For the case of this study, this function relates to the duration of unemployment. Typically, the most widely used estimator for this estimate is the Kaplan-Meier estimator, which allows for right-censored data. This estimator calculates the probability of knowing the event in each time interval, and we thus obtain a curve which is interpreted simply as the proportion of "survivors" for each length of stay in a given state. In other words, the proportions of individuals leaving unemployment for each period of unemployment.

In this application we will analyze the durations of global unemployment in the local employment agency of Ain El Benian. We are working on a sample of 1064 unemployed individuals observed between 01/01/2011 and 07/15/2013. By distinguishing those who found a job, 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 entered a job after his daily period of unemployment t_i.. This method finds the average chance risk value and compares it with the national unemployment rate, in this way it is possible to measure the insertion value of the unemployed at this agency.

2. Analysis of unemployment durations

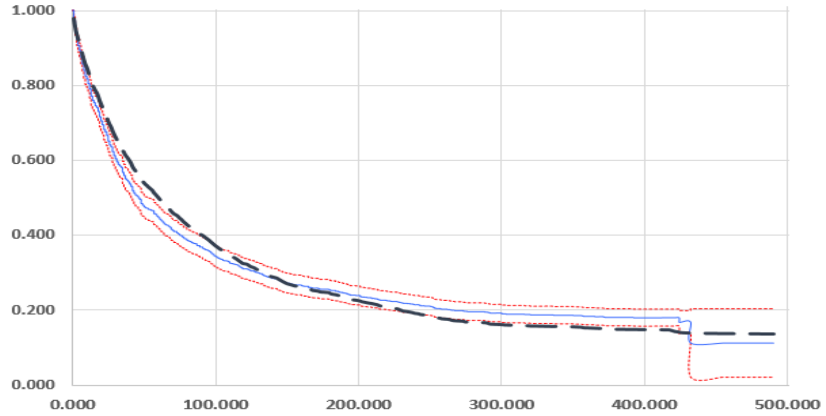
From Figure (2), we notice that at the start of the curve, 100% of the individuals in the sample are unemployed. After approximately 2 months of registering with this agency, 50% of individuals were placed in the labor market. But, the exit from unemployment for the rest of the individuals in the sample is spread out over a long period, for some it even exceeds a year. In general, from the unemployment duration curve, we deduce that the probability of exiting unemployment for those registered with 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 : Kaplan-Meier survival functions for overall unemployment duration.



Source: Developed by us using Excel.

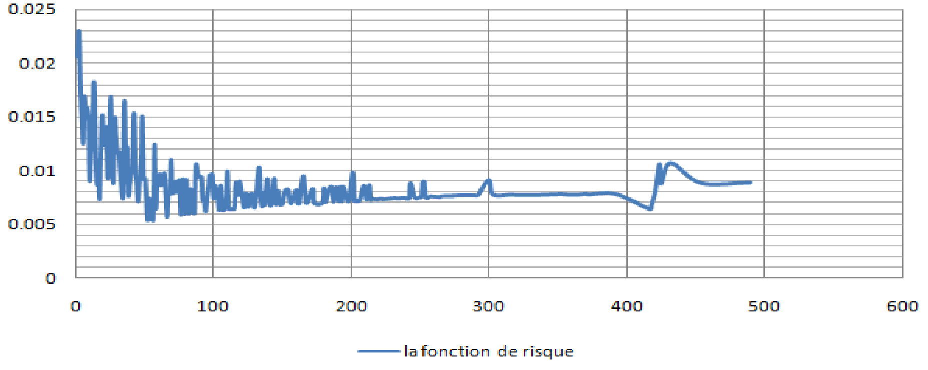
Figure 3: The evolution of survival probabilities in the classical method (in blue) of KM and Bayesian with a vague prior law (the discontinuous curve)



Source: Developed by us using Excel.

In Figure (3) we notice a small difference between the estimates of unemployment durations between the Bayesian and classical method, even at the median level (45 days). Bayesian curves represent smoother shapes compared to the frequentist or classical method. Consequently, the frequentist approach presents in this example (the Kaplan Meier model and the Bayesian model with a vague a priori of beta (0.01, 0.01)) a special case of Bayesian inference.

Figure 4 : the risk function



Source: Developed by us using Excel.

From Figure (4) we see that the probability of exiting unemployment gradually decreases. This leads us to deduce that there is a negative duration dependency. Initially, the probability of exiting unemployment is 0.02 for job seekers who experience short durations of unemployment. Then, the risk begins to gradually decrease until it reaches a very low probability of exiting unemployment of less than 0.009 for the unemployed with durations of more than 200 days.

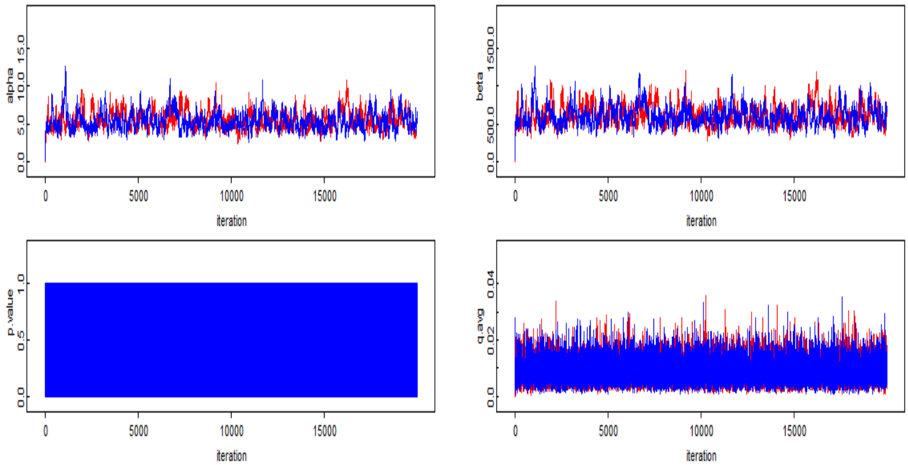
Tableau 2 : the parameters estimated by the unemployment study.

	mean	sd	MC_error	val2.5pc	median	val97.5pc
alpha	5.444	1.217	0.05482	3.549	5.255	8.245
beta	610.2	131.9	5.912	401.6	591.0	912.7
p. Value	0.6999	0.4583	0.003778	0.0	1.0	1.0
q.avg	0.008843	0.003885	2.067E-5	0.002918	0.008314	0.01795

Source: Developed by us using OpenBUGS.

The a posteriori predictive p-value (p-value = 0.69) can be directly interpreted as the probability of observing in future samples with $D(d_i, q_i{}^k)$ higher than that already observed. This value is close to 0.5 so the repeated and real data distributions are close, while values close to zero or one indicate differences between them (Gelman and Meng, 1996). The Hierarchical Bayesian Kaplan Meier model effectively represents the sample. The value of q.avg indicates that the average chance of dying during a day when taking prednisolone is 0.8%. the average national unemployment rate in Algeria between (2011 and 2013) is 10.27% according to the estimates of the International Monetary Fund, therefore the local employment agency of Ain El Benian between (2011 and 2013) allows to reduce the unemployment rate in a pretty big way. The chance of finding a job through integration increased by almost 13 times (12.83) among individuals registered with the Ain El Benian local employment agency between (2011 and 2013).

Figure5: The trace of the posterior distribution for the model parameters

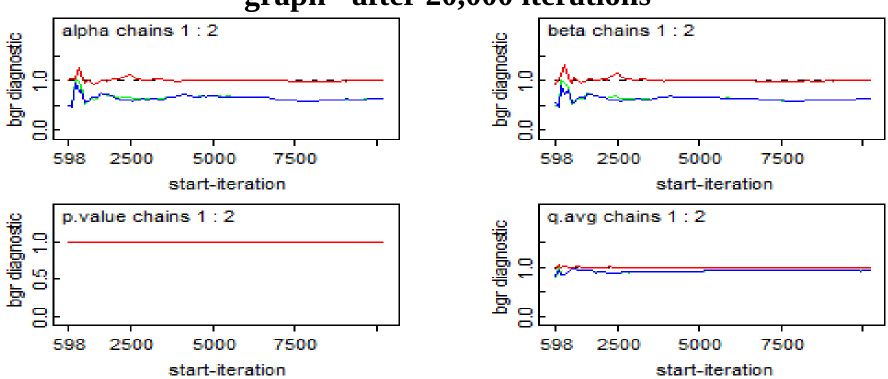


Source: Developed by us using OpenBUGS.

In Figure 5, each color denotes an MCMC chain. The two chains mix well: convergence is achieved (see A1 in the appendices).

Brooks and Gelman in 1998 proposed a generalization of the method of Gelman and Rubin which was introduced in the year 1992, it is a method of validating ergodic sequences of MCMC algorithms.

Figure 6: The Brooks and Gelman graph "convergence - diagnosis - graph" after 20,000 iterations



Source: Developed by us using OpenBUGS.

The green curve indicates the width of the 80% inter-chain credibility interval. The blue curve indicates the average width of the within-chain 80% credibility intervals. The red curve indicates the Brooks and Gelman statistic (i.e., the ratio of the green / blue curves). The Brooks and Gelman statistic tends towards 1, which means that there is convergence.

VI. Conclusion

When the intrinsic utility of hierarchical Bayesian modeling is to leverage the difficulty in calculating the a posteriori distribution and thanks to the flexibility of Bayesian models it is possible to combine these two magnificent solutions from a very well-known idea in non-parametric statistic of the duration model this is the Kaplan Meier model. Consequently, we calculate the average value of the risk of chance and by comparing it with the national unemployment rate, it is possible to measure the integration value of the unemployed at this agency. In the results we find:

• The chance of finding a job through integration increased by almost 13 times (12.83) among individuals registered with the Ain El Benian local employment agency between (2011 and 2013).

• This method allows comparisons to be made with great precision between employment agencies at the national level and between and within the wilayas.

VII. References

- ACHCAR, J.A., BOLFARINE, H., PERICCHI, L.R(1986), Transformation of survival data to an extreme value distribution, the statistician, 36, 229-234.
- ACHCAR, J.A., BROOKMEYER, R., HUNTER, W.G (1985), An application of Bayesian analysis to medical follow-up data, Statistics in Medicine, 4, 509-520.
- Boreux, J., Parent, E., Bernier, J (2010), Pratique du calcule Bayésien. Springer.

- Brooks, S. and Gelman, A (l998), “Alternative methods for monitoring convergence of iterative simulations”, Journal of Computational and Graphical Statistics 7,434455.
- CARLIN, B.P., GELFAND, A.E., SMITH, A.F.M (1992), Hierarchical Bayesian Analysis of Change point Problems, Appl. Statist. 41(2), 389-405.
- CHEN, W.C., HILL, B.M., GREENHOUSE, J.B., FAYOS, J.V (1985), Bayesian analysis of survival curves for cancer patients following treatment, In Bayesian Statistics 2 (Eds. J.O. Berger, J. Bernardo. A.F.M. Smith), Amsterdam: North-Holland, 299-328.
- 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.
- Cox, D. R (1972), Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2), 187-202.
- DELLAPORTAS, P., SMITH, A.F.M (1993), Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling, Applied Statistics, 42, 443-459.
- FERGUSON, T.S (1973), A Bayesian analysis of some nonparametric problems, The Annals of Statistics, 1(2), 209-230.
- FLORENS, J.P., ROLIN, J.M (2001), Simulation of posterior distributions in nonparametric censored analysis, International Statistical Review, 67(2), 187-210.
- Gelman, A. and Meng, X.-L (1996), “Model checking and model improvement”, in W. Gilks, S. Richardson, and D. Spiegelhalter, eds., Markov Chain Monte Carlo in Practice, Chapman & Hall, Suffolk, UK, pp. 189-201.
- GRIEVE, A.P(1987), Applications of Bayesian software: Two examples, The Statistician, 36, 283-288.
- IBRAHIM, J. G., CHEN, M, H., SINHA, D (2001), Bayesian survival analysis, Springer.

- KALBFLEISCH, J.D (1978), Nonparametric bayesian analysis of survival time data, Journal of the Royal Statistical Society, Series B, 40(2), 214-221.
- KALBFLEISCH, J.D., PRENTICE, R.I (1980), The Statistical analysis of failure time data, John Wiley & Sons, New York.
- KIM, S.W., IBRAHIM, J.G (2001), On Bayesian inference for proportional hazards models using noninformative priors, Lifetime Data Analysis, to appear.
- Meng, X.-L (l994), "Posterior predictive p-values", Annals ofstatistics 22, 1142-1 160.
- Robert, C.P (2006), Le choix Bayésien : principes et pratiques. Springer.

VIII. Appendices (OpenBUGS Code)

```
model
{
for (i in 1:m) {
d[i]~dbin(q[i],n[i])
q[i]~dbeta(alpha,beta)
d.rep[i] ~ dbin(q[i], n[i]) #Replicate from posterior predictive distribution
diff.obs[i] <- pow(d[i] - q[i]*n[i], 2)/(n[i]*q[i]*(1-q[i]))
diff.rep[i] <- pow(d.rep[i] - q[i]*n[i], 2)/(n[i]*q[i]*(1-q[i]))
}
for (i in 1:m){
ce[i]~dbin(qc[i],n[i])
}
for (i in 1:m){
qc[i]~dbeta(1,1)
}
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-1]*p[i]
}
s[1]<-p[1]
q.avg ~ dbeta(alpha, beta) # Average beta population variability curve
alpha ~ dgamma(0.0001, 0.0001)
beta ~ dgamma(0.0001, 0.0001)
chisq.obs <- sum(diff.obs[])
chisq.rep <- sum(diff.rep[])
p.value <- step(chisq.rep - chisq.obs)
}
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,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,36,39,29,
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(alpha=0.11, beta=0.22),
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.
5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,
0.5,0.5,0.5))
list(alpha=0.01, beta=0.01,
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))