Bayesian Constant hazard risk model with a change point
Case of Study: the durations of unemployment of a local employment agency

*** Hamimes Ahmed [1], Benamirouche Rachid [2]**

[1] PHD Student, National Higher School of statistics and applied economics (Algeria), ahmedhamimes@yahoo.com

[2] Professor, National Higher School of statistics and applied economics (Algeria), rbena2002@hotmail.com

Received: 08/11/2020 *Accepted: 12/12/2020* *Published: 31/12/2020*

Abstract:

The purpose of the change methods is to make statistical inferences about the position of breakpoints and about other model parameters. In this article we look at the unemployed registered with the local employment agency of Ain El Benian (January 2011-July 2013). The objective is to find the breaking point in the overall survival function represents the integration probabilities of individuals registered with this agency and in the determined period. We use the constant model of instantaneous risk which corresponds to an exponential function of survival. This breaking point represents the duration of change for an unemployed individual, which is an important element for economic analysis and comparison.

Key words: insertion probabilities, survival function, constant model of instantaneous risk, point of change.

Classification JEL: C25, C41, D81, E24.

*** Corresponding author.**

Introduction

The world is constantly changing. Being aware of these changes will help people avoid unnecessary losses and benefit from transitions in their favor. From a statistical point of view, the violation of the stochastic homogeneity of the data can be an indication of certain changes in the observed process (breakdowns, malfunctions, etc.). In fact, the statistician is faced with the problem of detecting the position and number of offset (break) points in statistical data, and in many realistic circumstances. In many areas, many of the practical problems of detecting change can be encountered. Change point problems arise in many areas, such as, signal processing, industrial control systems, economics, medicine, agronomy, meteorology, etc.

Survival analysis appeared in the twentieth century, and in the second half of the century, it experienced considerable development. Developments in this area, which have had the most profound impact on clinical trials, include the Kaplan-Meier (1958) method for estimating survival function, log-rank statistics (Mantel, 1966) for comparing two survival distributions, and the proportional hazard model (Cox, 1972) to quantify the effects of covariates on survival time. The martingale theory for the counting method, introduced by Aalen (1975), provides a consistent framework for studying the properties of statistics on survival analysis, for both small and large samples. Significant progress has been made and more improvement can be hoped for in many areas, such as the time-accelerated model, multivariate survival data, interval-censored data, complex treatment protocols, and causal inference. , modeling. Popular longitudinal data and survival data, and Bayesian methods.

In clinical trials, it may also be that one of the goals is to investigate the effect of a new treatment on patients' risk rates and whether it creates a substantial difference in risk rates. After the start of risk therapy continues. It would therefore be useful to define the breaking point of this risk rate, and thus to assess the period of time during which the treatment was very successful.

Matthews and Farewell (1982, 1985) were the first to address the issue of checking the risk function for a breakpoint, which is not an index of observations. They calculated the critical values of the Likelihood Ratio test for a breakpoint using simulation techniques. In addition to the work of Davies (1977), Matthews, Farewell and Pyke (1985) propose another alternative to check the breaking point in the instantaneous survival method based on the process of score statistics. He inspired Liang, Self and Liu (1990) to establish the Cox proportional hazards model for a particular case of time-dependent covariates, where they change their values only once over time, and use the score statistics to test the breaking point in the instantaneous risk function.

In this article we focus on the unemployed registered with the local employment agency of Ain El Benian (January 2011-July 2013). The objective is to find the breaking point in the overall survival function represents the integration probabilities of individuals registered with this agency and in the determined period. We use the constant model of instantaneous risk which corresponds to an exponential function of survival. This breaking point represents the duration of change for an unemployed individual which represents an important element for the period of high integration, the period of low integration and the critical point of the survival function, i.e. the point that requires special treatment by those in charge to increase the probability of inclusion. This method (the detection of break points by a constant chance model) also makes it possible to make comparisons between agencies in several wilayas or between regions.

I. Model with a change point

1. The change model

The simplest model of survival is the constant model of instantaneous risk which corresponds to an exponential function of survival. This model assumes that over time the risk of an event does not change. However, instantaneous changes in risk are crucial to detect. Simple change trends have the advantage of being easy to describe and understand. One of the simplest models of instantaneous change in risk is the constant risk model with a breaking point. Let T be the time for a specific case (a particular event), such as the time elapsed for patients with leukemia before relapse after diagnosis with "remission induction". Matthews and Farewell (1982) considered a model specified by the risk function for the distribution of T:

$$\lambda(t) = \begin{cases} \lambda & si\ t \leq k, \\ (1-\alpha)\lambda & si\ t > k \end{cases} \qquad (1)$$

This model has three parameters, λ, α ($0 \leq \alpha < 1$), and the break point k, which are all unknown. In other words, the hazard rate is constant up to an unknown point k, after which it takes on a new constant value.

The density and the survival function are respectively:

$$f_T(t) = \begin{cases} \lambda e^{\lambda t} & si\ t \leq k, \\ \rho\lambda e^{\lambda k - \rho\lambda(t-k)} & si\ t > k \end{cases} \qquad (2)$$

and

$$S_T(t) = \begin{cases} e^{\lambda t} & si\ t \leq k, \\ e^{\lambda k - \rho\lambda(t-k)} & si\ t > k \end{cases} \qquad (3)$$

where $\rho = 1 - \alpha$, and where λ, δ and k are the parameters of the model (1).

We consider $\varepsilon_i = 1$ si $T_i < k$ and $\varepsilon_i = 0$ si $T_i \geq k$, the likelihood function is:

$$L(\lambda,\rho,k) = \prod_{i=1}^{n} \left(\lambda e^{-\lambda t_i}\right)^{\varepsilon_i} \{[\rho\lambda exp\{-\lambda k - \rho\lambda(t_i - k)\}]^{1-\varepsilon_i}\}^{\delta_i}$$

$$* \prod_{i=1}^{n} (\lambda e^{-\lambda t_i})^{\varepsilon_i} \{[exp\{-\lambda k - \rho\lambda(t_i - k)\}]^{1-\varepsilon_i}\}^{1-\delta_i}$$

therefore

$$L(\lambda,\rho,k) = \lambda^{d_3}\rho^{d_3 - d_1(k)} exp\{-\lambda[S_1(k) + \rho S_2(k)]\} \qquad (4)$$

where

$$d_1(k) = \sum_{i=1}^{n} \delta_i\varepsilon_i, d_2(k) = \sum_{i=1}^{n} \varepsilon_{i,} d_3 = \sum_{i=1}^{n} \delta_i,$$

$$w_1(k) = \sum_{i=1}^{n} \delta_i\varepsilon_i t_i + \sum_{i=1}^{n} (1-\delta_i)\varepsilon_i t_i,$$

$$w_2(k) = \sum_{i=1}^{n} \delta_i (1-\varepsilon_i) t_i + \sum_{i=1}^{n} (1-\delta_i)(1-\varepsilon_i) t_i,$$

$$w_3(k) = \sum_{i=1}^{n} \delta_i (1-\varepsilon_i) + \sum_{i=1}^{n} (1-\delta_i)(1-\varepsilon_i) = n - d_2(k),$$

$$S_1(k) = w_1(k) + k w_3(k),$$

$$S_2(k) = w_2(k) - k w_3(k)$$

2. **The Bayesian representation of the model with a change point**

We suppose the change point k take discrete values $k_i = t_i$, with a prior distribution $\pi_0(k_i = t_i), i = 1,2, \ldots m,$ when m is the size of the sample. The a priori density of λ, ρ and of $k_i$is:

$$\pi(\lambda, \rho, k_i) = \pi(\lambda, \rho / k_i = t_i) \pi_0(k_i = t_i) \qquad (5)$$

Given $k_i = t_i$, assuming an approximate independence between the parameters λ and ρ, a priori uninformative density for λ and ρ is:

$$\pi(\lambda, \rho / k_i = t_i) \propto \frac{1}{\lambda \rho}, \quad \lambda, \rho > 0 \qquad (6)$$

The posterior density of λ, ρ and of k is:

$$\pi(\lambda, \rho, k / \mathcal{D}) \propto \lambda^{d_3 - 1} \rho^{d_3 - d_1(k) - 1} exp\{-\lambda [S_1(k) + \rho S_2(k)]\} \qquad (7)$$

such that $\mathcal{D}$ represents the set of data.

The a posteriori conditional densities for the Gibbs algorithm are given by:

$$\pi(\lambda / \mathcal{D}, \rho, k) \sim Gamma(d_3, S_1(k) + \rho S_2(k))$$

$$\pi(k / \mathcal{D}, \rho, \lambda) \propto \rho^{-d_1(k)} exp\{-\lambda [S_1(k) + \rho S_2(k)]\}$$

$$\pi(\rho / \mathcal{D}, \lambda, k) \propto \sum_{i=1}^{T} \int_0^{+\infty} \int_0 \lambda^{d_3 - 1} \rho^{d_3 - d_1(k) - 1} exp\{-\lambda [S_1(k) + \rho S_2(k)]\} d\mathcal{D} d\lambda dk$$

II. Description and presentation of statistical data

Faced with the worsening of the unemployment phenomenon, programs to combat unemployment have been implemented. We can classify the solutions recommended by the public authorities into two categories: passive solutions or social treatment of unemployment and active solutions1 or real job creation.

1. **The National Employment Agency**

Pour tous les demandeurs d'emploi pour la première fois (ceux qui n'ont jamais travaillé) et pour tous ceux qui sont au chômage et à la recherche d'un emploi, l'Agence nationale pour l'emploi est un must. C'est le premier pas pour un étudiant diplômé ou un stagiaire qui a terminé la préparation. Pour bénéficier des mécanismes d'intégration, vous devez être enregistré auprès de l'ANEM. Cette agence a également besoin que les entreprises publiques et privées passent par ses services de recrutement et nomment tous les nouveaux travailleurs. Le système d'assistance technique à l'intégration (DAIP) est en pleine révision et modification depuis 2008. L'ancien système de pré-emploi n'a pas donné les résultats escomptés. Dans la plupart des cas, il a produit des emplois précaires qui se traduisent rarement par un recrutement permanent.

Compte tenu de la nouveauté de ces dispositifs, aussi obligatoire que soit le passage de tout recrutement par l'Agence nationale pour l'emploi, on peut s'attendre à une nette augmentation du taux de placement de cette organisation, ce qui devrait rendre l'action publique plus visible dans ce domaine.

2. Local Employment Agency: (ALEM)

The local agency is considered to be annexes of the wilayas marked by high concentrations of population and activity. The ALEM is the last stage of the company, introduced at the municipal level, where they will specialize in prospecting for donations. However, they must direct their activities to the resident communities.

a. Data presentation

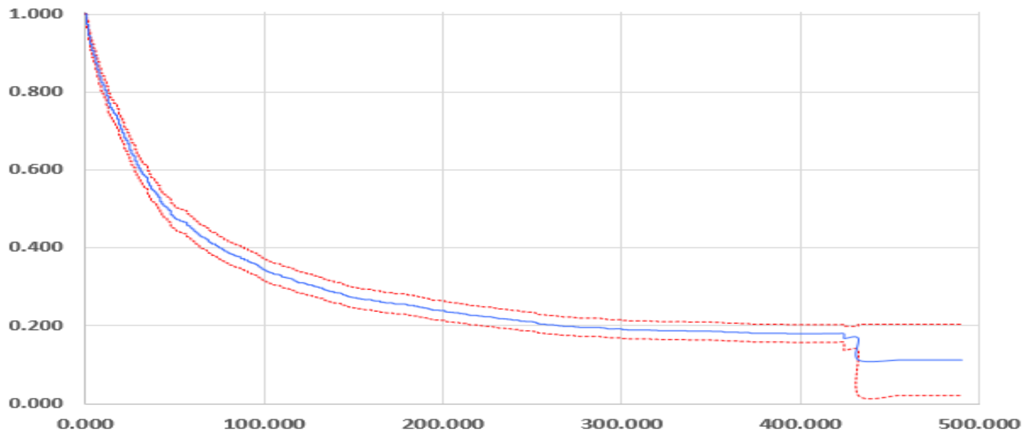
One of the most effective tools for analyzing lifespans is certainly the survival or stay function estimator. For the case of this study, this function relates to the duration of unemployment. Typically, the most widely used estimator for this estimate is the Kaplan-Meier estimator, which allows for 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 duration 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.

b. Analysis of unemployment durations

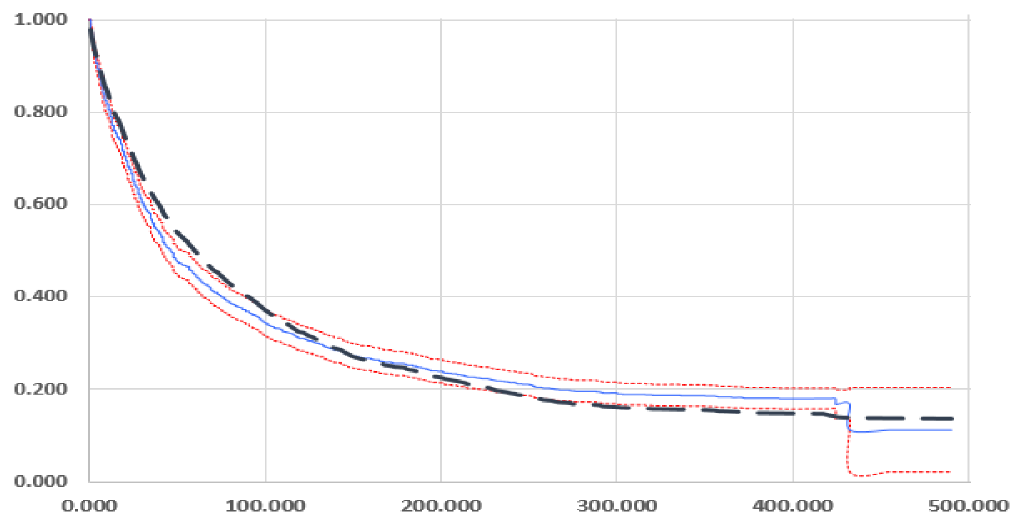
From Figure (1), we notice that at the start of the curve, 100% of the individuals in the sample are unemployed. After approximately 2 months of registering with this agency, 50% of individuals were placed in the labor market. But, the exit from unemployment for the rest of the individuals in the sample 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.

Fig n°(0 1) : Kaplan-Meier survival functions for overall unemployment duration.



Source: Developed by us, using Excel program.

Fig n°(2) : 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 program.

In Figure (2), 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 particular case of Bayesian inference.

c. Change point analysis

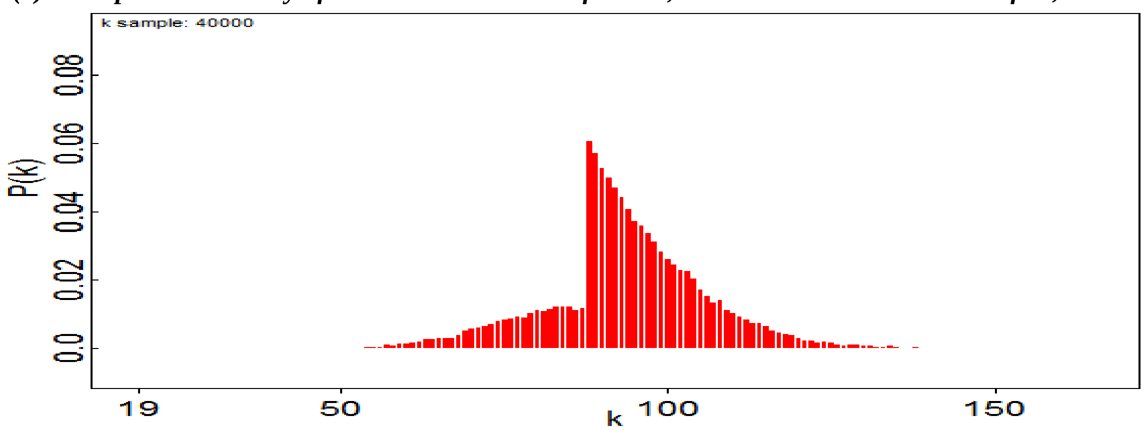
According to table (1) the change in the insertion probability is 20% in the found change point $k \cong 94$, According to the program code WinBUGS (see appendix) assuming that $k = N(x_k)$,, so the change point is $\tau_1 \cong 97$ days (see figure 2). So after almost 100 days of registration at the local employment agency of Ain El Benian (January 2011-July 2013) we notice a change in the insertion rate from a period of high integration to a period of low insertion. The point of change in the studied series requires special treatment to increase the probability of insertion.

Table n°(0 1) : the parameters estimated by the study of change in unemployment.

	mean	sd	MC_error	val2.5pc	median	val97.5pc
alpha	0.7889	0.03347	4.42E-4	0.7176	0.7909	0.848
k	93.68	12.08	0.1661	68.0	93.0	119.0
lambda[1]	0.02146	0.003748	6.058E-5	0.01532	0.02106	0.03027
lambda[2]	0.004434	4.672E-4	3.546E-6	0.003547	0.004426	0.005387

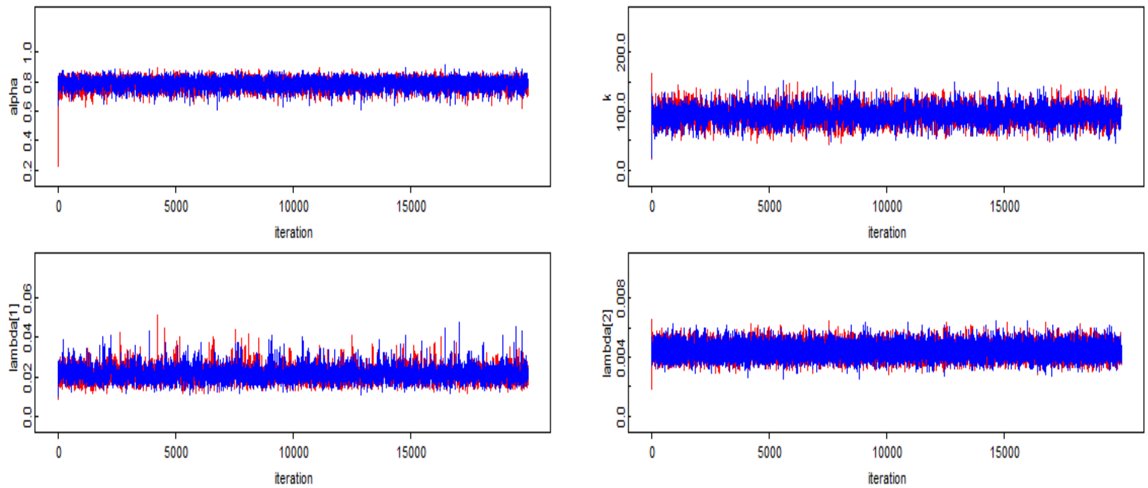
Source: Developed by us, using OpenBUGS program.

Fig n°(3) : The posterior density of $\pi(k / x, \lambda_1, \lambda_2, \alpha)$ ***after 200,000 iterations with a burn-in of 10,000 iterations.***



Source: Developed by us, using OpenBUGS program.

Fig n°(0 4) : The trace of the posterior distribution for the model parameters.

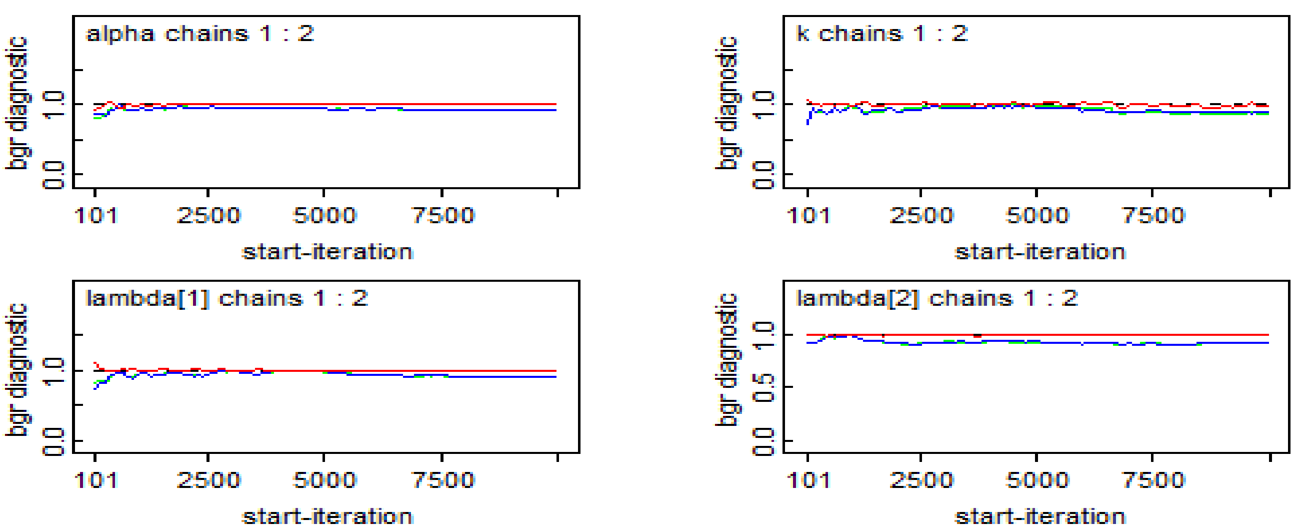


Source: Developed by us, using OpenBUGS program.

In Figure 4, 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.

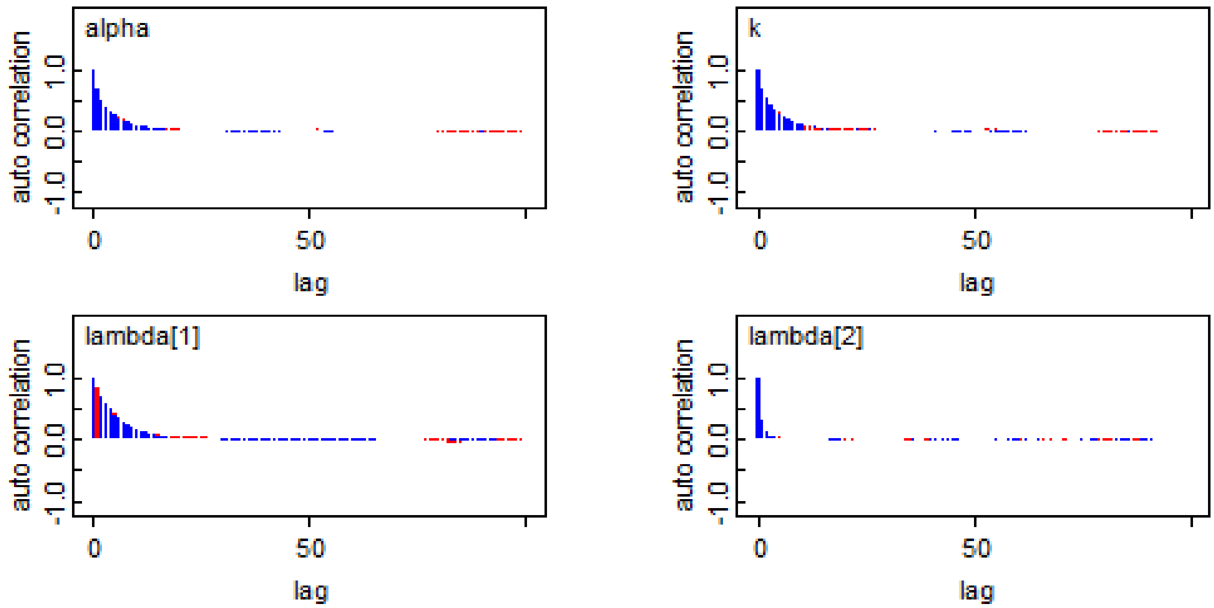
Fig n°(05) : The Brooks and Gelman graph "convergence - diagnosis - graph" after 20,000 iterations.



Source: Developed by us, using OpenBUGS program.

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.

Fig n°(06) : The autocorrelation graph after 20,000 iterations in the model studied.



Source: Developed by us, using OpenBUGS program.

The cause of autocorrelation is that the parameters of our model can be highly correlated, so the Gibbs sampler will be slow to explore the entire distribution a posteriori. If the level of autocorrelation is high for a parameter of interest, then a trace plot will be a poor diagnosis for convergence. In our study we observe a weak autocorrelation for the parameters of interest, so this is another sign of good convergence.

IV- Conclusion

The analysis by the_breaking point (change), shows that after almost 100 days of registration at the local employment agency of Ain El Benian (January 2011-July 2013) we notice a change in the rate of 'insertion from a period of high integration into a period of low integration. This means that the unemployed registered the first 97 days they have the same probability of integration, and the reverse for the unemployed registered in a durations exceeding 97 days. This point of change is a reality to compare the efficiency of agencies, the more the duration of high insertion increases the more the efficiency of the agency increases. The same goes for the parameter (1-alpha) which represents the percentage decrease in insertion rate at this agency after the break point. For an employment agency, it is important to increase the duration of high integration and the deprived parameter (1-alpha).

Appendices (the OpenBUGS code)

```
model {
for(i in 1:m) {
t[i] ~ dexp(mu[i]) I(c[i], )
mu[i]<- lambda[J[i]]
J[i]  <- 1 + step(i - (k+0.5))
p[i] <- 1/m
}
 lambda[1]~ dgamma(0.01,0.01)
 lambda[2]<-(1-alpha)* lambda[1]
alpha~ dbeta(1,1)
k ~ dcat(p[])
}
list(k=32,  alpha=0.2, lambda=c(0.2, NA))
list(k=137,  alpha=0.23, lambda=c(1.2, NA))
DATA
list( m = 227,
t=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,
31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,
59,60,61,62,63,64,65,66,67,68,69,70,71,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,89,1,92,93,9
4,95,96,97,98,99,100,101,102,103,104,105,106,108,109,110,111,112,113,114,115,116,117,118,119,
122,123,124,125,127,128,129,130,131,133,134,135,136,137,139,141,143,144,145,146,147,148,149,
150,152,154,156,160,161,162,165,167,168,172,173,179,180,181,184,186,187,189,191,192,193,194,
198,201,202,204,206,209,211,213,214,215,219,220,222,
223,227,228,229,230,231,234,236,238,239,242,243,246,250,251,252,253,254,257,258,262,
264,265,267,273,274,276,288,290,291,292,300,303,316,320,336,350,356,357,365,371,372,
392,NA,NA,NA,NA,NA,424,NA,432,NA,NA),
c=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,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,417,418,419,420,421,0,425,0,454,490))
```

References

1. Aalen, O. O. (1975). Statistical inference for a family of counting process (Doctoral dissertation, Ph. D. thesis, University of California, Berkeley).
2. Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2), 187-202.
3. Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika, 64(2), 247-254.
4. Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282), 457-481.

5. Liang, K. Y., Self, S. G., & Liu, X. (1990). The Cox proportional hazards model with change point: An epidemiologic application. Biometrics, 783-793.
6. Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother. Rep., 50, 163-170.
7. Matthews, D. E., & Farewell, V. T. (1982). On testing for a constant hazard against a change-point alternative. Biometrics, 463-468.
8. Matthews, D. E., Farewell, V. T., & Pyke, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. The Annals of Statistics, 13(2), 583-591.
9. Matthews, D. E., Farewell, V. T., & Pyke, R. (1985). Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative. The Annals of Statistics, 13(2), 583-591.