

Note on the Linearity of Bayesian Estimates in the Dependent Case

Souad Assoudou¹, Belkheir Essebbar²

¹Department of Economics, Faculty of Law, Economics and Social Sciences, Hassan I University, Settat, Morocco

²Department of Mathematics and Computer Sciences, Faculty of Science, Mohammed V University, Rabat, Morocco

Email: s_assoudou@yahoo.fr

Received September 21, 2013; revised October 21, 2013; accepted October 28, 2013

Copyright © 2014 Souad Assoudou, Belkheir Essebbar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for SCIRP and the owner of the intellectual property Souad Assoudou, Belkheir Essebbar. All Copyright © 2014 are guarded by law and by SCIRP as a guardian.

ABSTRACT

This work deals with the relationship between the Bayesian and the maximum likelihood estimators in case of dependent observations. In case of Markov chains, we show that the Bayesian estimator of the transition probabilities is a linear function of the maximum likelihood estimator (MLE).

KEYWORDS

Bayes Estimator; Maximum Likelihood Estimator; Markov Chain; Transition Probabilities; Jeffreys' Prior; Multivariate Beta Prior; MCMC

1. Introduction

Let X be the random variable with Bernoulli distribution with parameter θ . It's known that the Bayesian solution, under quadratic loss and a beta distribution $Be(\alpha, \beta)$ as prior for θ , is given by

$$E(\theta|X) = aX + b, \quad (1.1)$$

with $a = 1/(\alpha + \beta + 1)$ and $b = \alpha/(\alpha + \beta + 1)$. If X_1, \dots, X_n is a sample from X , the Bayesian estimate becomes

$$E(\theta|X_1, \dots, X_n) = a_n \bar{X} + b_n, \quad (1.2)$$

with $a_n = n/(\alpha + \beta + n)$ and $b_n = \alpha/(\alpha + \beta + n)$.

Moreover, \bar{X} is the MLE which coincides with the empirical estimate.

Formulas (1.1) and (1.2) are still true with distribution of X in the exponential family [1].

Let's now move to the dependent case and show that (1.2) is still true for the Markov chains.

Let $X = (X_0, \dots, X_n)$ be the $n+1$ first observations of an homogeneous Markov chain with a finite state space $E = \{1, \dots, s\}$ and let $P = (p_{ij})$ be the matrix of transition probabilities. The likelihood is

$$L(x_0, \dots, x_n) = \prod_{t=1}^n L(x_t|x_{t-1}) \cdot P(X_0 = x_0), \quad (1.3)$$

where $L(j|i) = p_{ij}, i, j \in E$.

Let $N_n^{ij} = \sum_{t=1}^n 1_{\{X_{t-1}=i, X_t=j\}}$, which gives the number of one step transition from i to j until time n , then (1.3) becomes

$$L(x_0, \dots, x_n) = P(X_0 = x_0) \prod_{i,j} p_{ij}^{N_n^{ij}}. \tag{1.4}$$

The parameter $\theta = (p_{ij}, (i,j) \in E^2, i \neq j) \in]0, 1[^{s(s-1)}$ is to be estimated, and we assume that the distribution of X_0 is known, that is $\nu = (\nu_1, \dots, \nu_s)$ with $\nu_i = P(X_0 = i)$ is known. Then, the MLE of θ is given by

$$\hat{p}_{ij} = \frac{N_n^{ij}}{\sum_{j \in E} N_n^{ij}}, (i, j) \in E^2, \tag{1.5}$$

which coincides with the empirical estimate.

The Bayesian estimator under the quadratic error loss, is given by

$$\tilde{p}_{ij} = E(p_{ij} | X) = \int_0^1 p_{ij} \pi(p_{ij} | X) dp_{ij}, (i, j) \in E^2, \tag{1.6}$$

where $\pi(p_{ij} | X)$ is the posterior distribution. Two conjugate prior distributions are viewed. The multivariate beta distribution [2] and the Jeffreys' distribution [3].

This work is organized as follows: In the second Section, we develop the Bayesian estimation for different priors. In Section 3, the cases of 2 and 3 states Markov chain are discussed. A numerical study is given in Section 4.

2. The Bayesian Framework

In this section, we introduce two conjugate prior distributions for θ of a multistate Markov chain. The first kind of these priors is the product of the multivariate beta which doesn't take account of a possible correlation between the rows of transition probability matrix. The second kind is the Jeffreys' prior [3] which incorporates a certain type of dependence between the components of the parameter.

2.1. Multivariate Beta Priors

Given the characteristics of the transition probabilities p_{ij} , the natural choice of prior distributions is the multivariate beta distribution [2]. Thus, let the prior distribution for the elements of the i^{th} row of the matrix P be the multivariate beta, that is,

$$f(p_{i1}, \dots, p_{is}) = \frac{\Gamma\left(\sum_{j=1}^s a_{ij}\right)}{\prod_{j=1}^s \Gamma(a_{ij})} \prod_{j=1}^s p_{ij}^{a_{ij}-1}, \tag{1.7}$$

where Γ denotes the gamma function.

There are s rows in the transition matrix P and thus there are s such prior's in (1.7). If the rows of P are assumed independent then the joint prior probability density function for all p_{ij} is

$$\pi(\theta) = \prod_{i=1}^s f(p_{i1}, \dots, p_{is}) = \prod_{i=1}^s \frac{\Gamma\left(\sum_{j=1}^s a_{ij}\right)}{\prod_{j=1}^s \Gamma(a_{ij})} \prod_{j=1}^s p_{ij}^{a_{ij}-1}. \tag{1.8}$$

Making use of the likelihood given by (1.4) and the prior in (1.8), the joint posterior is

$$\pi(\theta | x) = \prod_{i=1}^s \frac{\Gamma\left(\sum_{j=1}^s (N_n^{ij} + a_{ij})\right)}{\prod_{j=1}^s \Gamma(N_n^{ij} + a_{ij})} \prod_{j=1}^s p_{ij}^{N_n^{ij} + a_{ij} - 1}.$$

The Bayesian estimator of p_{ij} corresponding to the multivariate beta is then given by

$$\bar{p}_{ij} = \frac{N_n^{ij} + a_{ij}}{\sum_{j \in E} (N_n^{ij} + a_{ij})}. \quad (1.9)$$

Let us remark that the Bayesian estimator \bar{p}_{ij} can be written as

$$\bar{p}_{ij} = A(N_n^{ij}) \hat{p}_{ij} + B(N_n^{ij}),$$

with $A(N_n^{ij}) = \sum_{j \in E} N_n^{ij} / (\sum_{j \in E} (N_n^{ij} + a_{ij}))$ and $B(N_n^{ij}) = a_{ij} / (\sum_{j \in E} (N_n^{ij} + a_{ij}))$. The above relation seems to be compared to (1.2), but here, the coefficients $A(N_n^{ij})$ and $B(N_n^{ij})$ depend also on the sufficient statistics N_n^{ij} .

2.2. The Jeffreys' Prior

Assoudou and Essebbar [3] have studied the Bayesian estimation for the multistate Markov chains, under the Jeffreys' prior distribution. As shown, this prior has many advantages: it permits a certain type of dependence between the various transition probabilities. Moreover it's a conjugate prior for $L(x_0, \dots, x_n)$ given by (1.4). The absence of extra parameter in this prior is of great interest, because we don't need to do more extra estimation.

The Jeffreys' prior and the correspondent posterior distributions [3] are respectively given by

$$\pi(\theta) \propto \left[\prod_{i=1}^s (v_i + (n-1)p_i) \right]^{\frac{s-1}{2}} \prod_{(i,j) \in E^2} p_{ij}^{\frac{1}{2}},$$

$$\pi(\theta|x) \propto \left[\prod_{i=1}^s (v_i + (n-1)p_i) \right]^{\frac{s-1}{2}} \prod_{(i,j) \in E^2} p_{ij}^{N_n^{ij} - \frac{1}{2}},$$

where $p = (p_1, \dots, p_s)$ is the stationary probability and v_i is the probability that the initial state X_0 is i .

One can deduce the Jeffreys' prior and posterior distributions for the two-states Markov chain,

$$\pi_1(\theta) \propto \left[\prod_{i=1}^2 (v_i + (n-1)p_i) \right]^{\frac{1}{2}} \prod_{(i,j) \in E^2} p_{ij}^{\frac{1}{2}}, \quad (1.10)$$

$$\pi_1(\theta|x) \propto \left[\prod_{i=1}^2 (v_i + (n-1)p_i) \right]^{\frac{1}{2}} \prod_{(i,j) \in E^2} p_{ij}^{N_n^{ij} - \frac{1}{2}}, \quad (1.11)$$

where $\theta = (p_{12}, p_{21})$ and $p = (p_1, p_2)$ is the stationary probability given by

$$p_1 = \frac{p_{21}}{p_{12} + p_{21}} \quad \text{and} \quad p_2 = \frac{p_{12}}{p_{12} + p_{21}}.$$

For the three-states Markov chain with parameter $\theta = (p_{12}, p_{13}, p_{21}, p_{23}, p_{31}, p_{32}) \in]0, 1[^6$, the prior and the posterior Jeffreys' distributions are respectively

$$\pi_2(\theta) \propto \left[\prod_{i=1}^3 (v_i + (n-1)p_i) \right] \prod_{(i,j) \in E^2} p_{ij}^{\frac{1}{2}}, \quad (1.12)$$

$$\pi_2(\theta|x) \propto \left[\prod_{i=1}^3 (v_i + (n-1)p_i) \right] \prod_{(i,j) \in E^2} p_{ij}^{N_n^{ij} - \frac{1}{2}}, \quad (1.13)$$

where the stationary probability $p = (p_1, p_2, p_3)$ is given by

$$p_1 = \frac{p_{21}p_{32} + p_{31}(p_{21} + p_{23})}{(p_{12} + p_{13})(p_{32} + p_{23}) + p_{31}(p_{12} + p_{21} + p_{23}) + p_{21}(p_{13} + p_{32})},$$

$$p_2 = \frac{p_{12}p_{31} + p_{32}(p_{12} + p_{13})}{(p_{12} + p_{13})(p_{32} + p_{23}) + p_{31}(p_{12} + p_{21} + p_{23}) + p_{21}(p_{13} + p_{32})},$$

$$p_3 = \frac{p_{13}p_{21} + p_{23}(p_{12} + p_{13})}{(p_{12} + p_{13})(p_{32} + p_{23}) + p_{31}(p_{12} + p_{21} + p_{23}) + p_{21}(p_{13} + p_{32})}.$$

3. The Bayesian Solution

For either posterior distributions (1.11) or (1.13), the integral given by (1.6) is difficult to calculate, so we propose an approximation of it by mean of an algorithm, namely the Independent Metropolis-Hasting algorithm (IMH) [4].

The fundamental idea behind these algorithms is to construct an homogeneous and ergodic Markov chain $(\theta^{(l)})$ with stationary measure $\pi(\theta|x)$.

For m “large enough”, $\theta^{(m)}$ is roughly distributed from $\pi(\theta|x)$ and the sample $\theta^{(m)}, \theta^{(m+1)}, \dots$, can be used to derive the posterior means. For instance, the Ergodic Theorem [4] justifies the approximation of the integral (1.6) by the empirical average

$$\tilde{\theta} = \frac{1}{m} \sum_{l=1}^m \theta^{(l)}$$

in the way that $\tilde{\theta}$ is converging to the integral (1.6) for almost every realization of the chain $(\theta^{(l)})$ under minimal conditions.

For next, we will give the description of this algorithm in the cases of two-states and three-states.

1) Case of the two-states Markov chain

Given $\theta^{(l)} = (\theta_1^{(l)}, \theta_2^{(l)})$

- Step1. Generate

$$y^{(l)} = (y_1^{(l)}, y_2^{(l)}) \sim U_{[0,1]} \times U_{[0,1]}.$$

- Step2. Take

$$\theta^{(l+1)} = \begin{cases} y^{(l)} & \text{with probability } \sigma_1(\theta^{(l)}, y^{(l)}) \\ \theta^{(l)} & \text{with probability } 1 - \sigma_1(\theta^{(l)}, y^{(l)}) \end{cases}$$

where

- $\sigma_1(\theta^{(l)}, y^{(l)}) = \min \left(\frac{\pi_1(y^{(l)}|x)}{\pi_1(\theta^{(l)}|x)}, 1 \right),$

- $\pi_1(\cdot|x)$ is the posterior density given by (1.11).

2) Case of the three-states Markov chain

Given $\theta^{(l)} = (\theta_1^{(l)}, \theta_2^{(l)}, \dots, \theta_6^{(l)})$

- Step1. Generate $y^{(l)} = (y_1^{(l)}, y_2^{(l)}, \dots, y_6^{(l)})$ such as:

1) Generate $(y_1^{(l)}, y_2^{(l)}) \sim U_{[0,1]} \times U_{[0,1]}$ until $y_1^{(l)} + y_2^{(l)} < 1$,

2) Generate $(y_3^{(l)}, y_4^{(l)}) \sim U_{[0,1]} \times U_{[0,1]}$ until $y_3^{(l)} + y_4^{(l)} < 1$,

3) Generate $(y_5^{(l)}, y_6^{(l)}) \sim U_{[0,1]} \times U_{[0,1]}$ until $y_5^{(l)} + y_6^{(l)} < 1$,

- Step2. Take

$$\theta^{(l+1)} = \begin{cases} y^{(l)} & \text{with probability } \sigma_2(\theta^{(l)}, y^{(l)}), \\ \theta^{(l)} & \text{with probability } 1 - \sigma_2(\theta^{(l)}, y^{(l)}), \end{cases}$$

where

$$\sigma_2(\theta^{(l)}, y^{(l)}) = \min \left(\frac{\pi_2(y^{(l)}|x)}{\pi_2(\theta^{(l)}|x)}, 1 \right),$$

- $\pi_2(\cdot|x)$ is the posterior density given by (1.13).

4. Numerical Study

In order to characterize the relationship between the Bayesian estimator and the MLE, it's indispensable to perform simulation studies. The first part of the analysis is devoted to Bayesian model founded on the multivariate beta prior given by (1.8) while the second deals with the Bayesian model using the Jeffreys' prior given by theorem 1. We discuss through this analysis the cases of Markov chain with two and three states.

In both experiments, a pascal program is written to run the transition probabilities. The Bayesian estimator corresponding to Jeffreys' prior is obtained from a single chain including 10^5 iterations of IMH algorithm (Section 3).

4.1. Experiment (a)

In Section 2.1, we have shown that, under the multivariate beta prior, the analytical Bayesian solution doesn't express clearly a certain linearity of the MLE. In this experiment, we demonstrate by simulation that the Bayesian estimator given by (1.9) is a linear function of the MLE given by (1.5).

4.1.1. Case of the Two-States Markov Chain

The data set is composed of 100 independent two-states Markov chains with 20 observations. To generate this data set, transition probabilities for each chain are drawn from the beta prior given by (1.8) with $s = 2$ and parameters $a_{ij} = 1/2$. The first state X_0 , in each chain, is assumed to be equal to 1 ($\nu_1 = 1$). The remaining observations given each chain, are then simulated in succession according to the appropriate transition probabilities. The Bayesian estimator \bar{p}_{ij} for each chain, is computed from (1.9).

Figure 1(a) (resp. **Figure 2(a)**) shows the plot of \bar{p}_{12} (resp. \bar{p}_{21}) versus \hat{p}_{12} (resp. \hat{p}_{21}). These figures prove that the Bayesian estimator \bar{p}_{ij} is quite a linear function of the MLE \hat{p}_{ij} .

4.1.2. Case of the Three-States Markov Chain

In this experiment we generate 100 independent Markov chains each with three states and 60 observations. By using the IMH algorithm, the transition probabilities are simulated from the multivariate beta given by (1.8) with $s = 3$ and parameters $a_{ij} = 1/2$. As previously, the Bayesian estimator \bar{p}_{ij} is calculated from (1.9). The plots of \bar{p}_{ij} versus \hat{p}_{ij} are given by the next figures (**Figure 3(a)** until **Figure 8(a)**). It's true that the linearity of the Bayesian estimates is still verified, but we remark that the plot is not quite a line as soon as the number of states grows.

4.2. Experiment (b)

Now let us consider the Bayesian model based on the Jeffreys' prior such that described in Section 2.2.

4.2.1. Case of the Two-States Markov Chain

In this experiment, we simulate 100 independent two-states Markov chains with $n = 20$, obviously the chains may be of differing lengths. The IMH algorithm is first used to generate the transition probabilities from the Jeffreys' prior given by (1.10). We assume, without loss of generality, that the first state X_0 , in each chain, is equal to 1 ($\nu_1 = 1$). The remaining observations, in each chain, are then simulated in succession with the appropriate transition probabilities. To obtain the Bayesian estimator \tilde{p}_{ij} of the transition probabilities given each chain, we apply the IMH algorithm to the posterior distribution given by (1.11). The MLE \hat{p}_{ij} is calculated from (1.5).

Figure 1(b) (resp. **Figure 2(b)**) displays the plot of \tilde{p}_{12} (resp. \tilde{p}_{21}) versus \hat{p}_{12} (resp. \hat{p}_{21}). From these figures it follows clearly that the Bayesian estimator \tilde{p}_{ij} is a linear function of the MLE \hat{p}_{ij} .

4.2.2. Case of the Three-States Markov Chain

The same experiment is repeated once more, but now the transition probabilities are drawn from the Jeffreys' prior given by (1.12) in order to generate 100 independent three-states Markov chains with $n = 60$. The Bayesian estimator is approximated by mean of IMH algorithm applied to the posterior distribution given by (1.13).

The next figures (Figure 3(b) until Figure 8(b)) show the plots of \tilde{p}_{ij} versus \hat{p}_{ij} . The same results as in Section 4.1.2 are obtained about the relationship between the Bayesian estimator and the MLE.

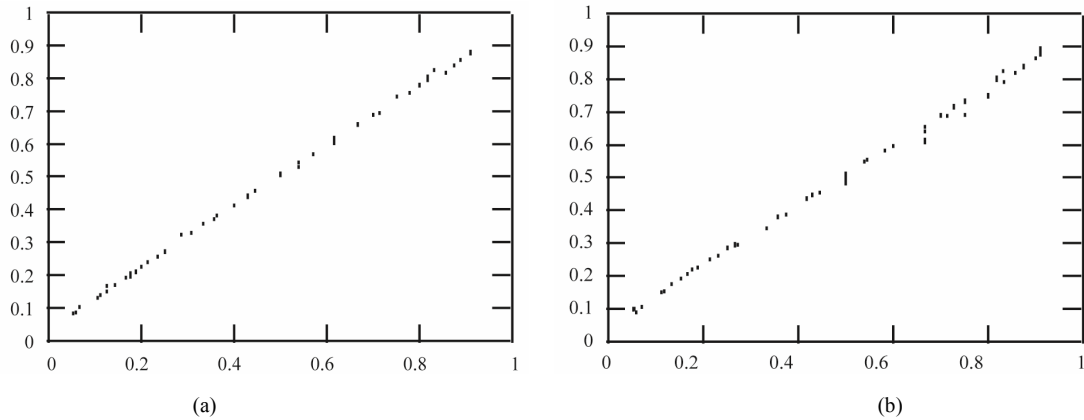


Figure 1. (a) Plot of \bar{p}_{12} versus \hat{p}_{12} ; (b) Plot of \tilde{p}_{12} versus \hat{p}_{12} .

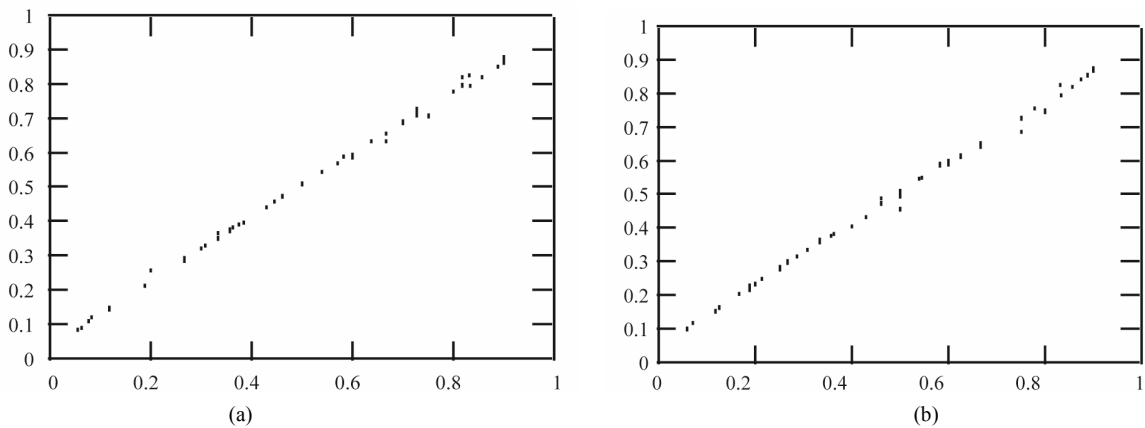


Figure 2. (a) Plot of \bar{p}_{21} versus \hat{p}_{21} ; (b) Plot of \tilde{p}_{21} versus \hat{p}_{21} .

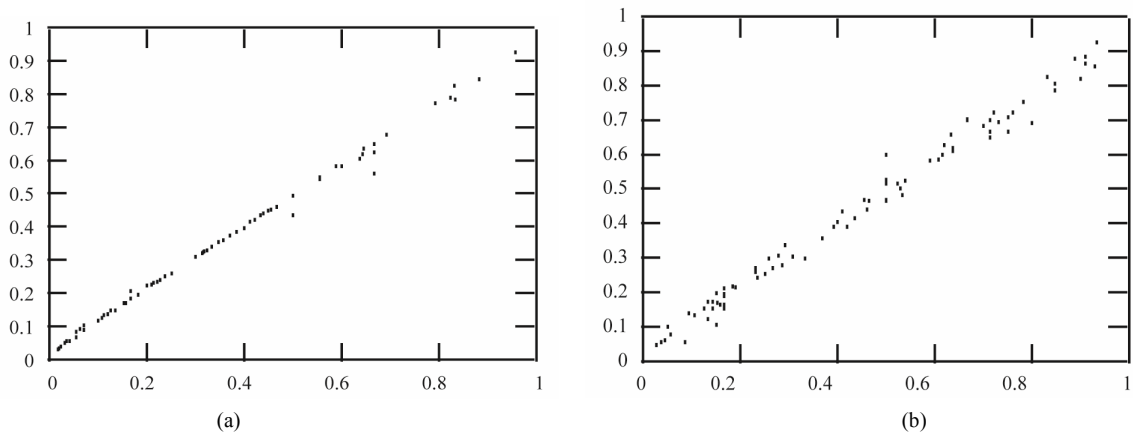


Figure 3. Plot of \bar{p}_{12} versus \hat{p}_{12} ; (b) Plot of \tilde{p}_{12} versus \hat{p}_{12} .

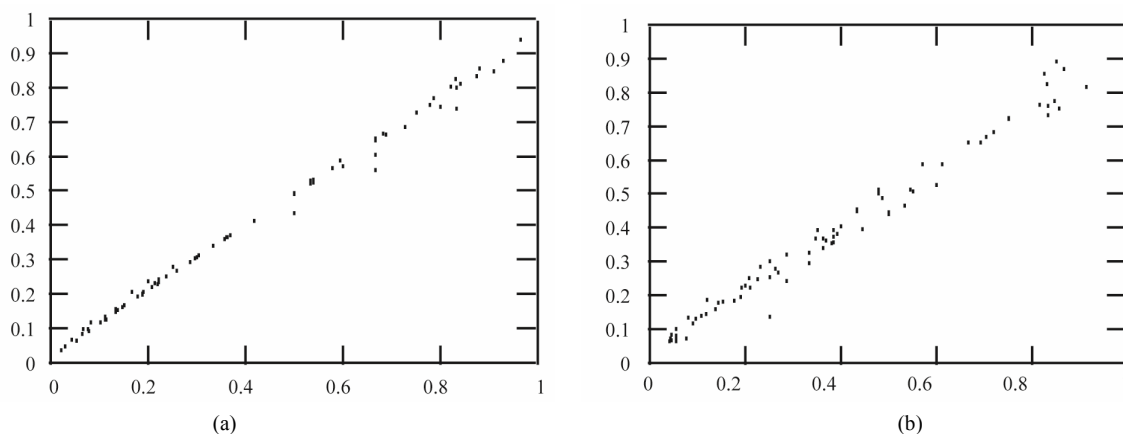


Figure 4. (a) Plot of \bar{p}_{13} versus \hat{p}_{13} ; (b) Plot of \tilde{p}_{13} versus \hat{p}_{13} .

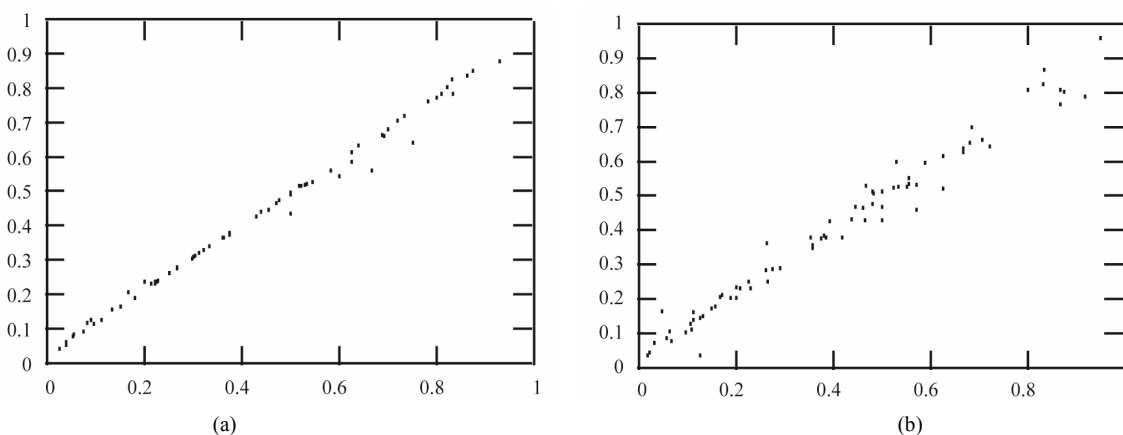


Figure 5. (a) Plot of \bar{p}_{21} versus \hat{p}_{21} ; (b) \tilde{p}_{21} versus \hat{p}_{21} .

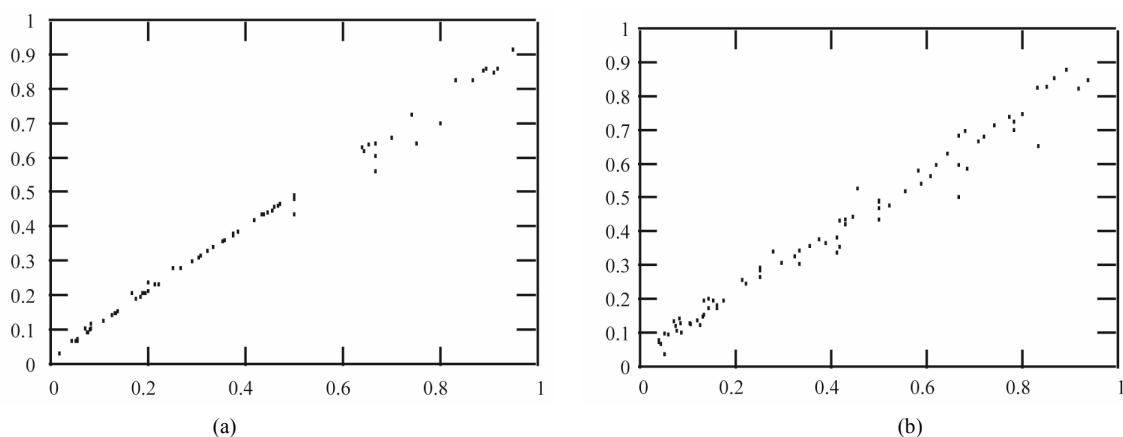


Figure 6. (a) Plot of \bar{p}_{23} versus \hat{p}_{23} ; (b) Plot of \tilde{p}_{23} versus \hat{p}_{23} .

5. Conclusion

The objective of this work is to study the relationship between the Bayesian estimator and the MLE in a dependent case such as a Markov chain. A numerical study by simulation is carried out to describe the nature of this relationship. Under the multivariate beta and the Jeffreys' priors we have shown that the Bayesian solution is still a linear function of the MLE. This linearity is now verified by simulation for these two models, others

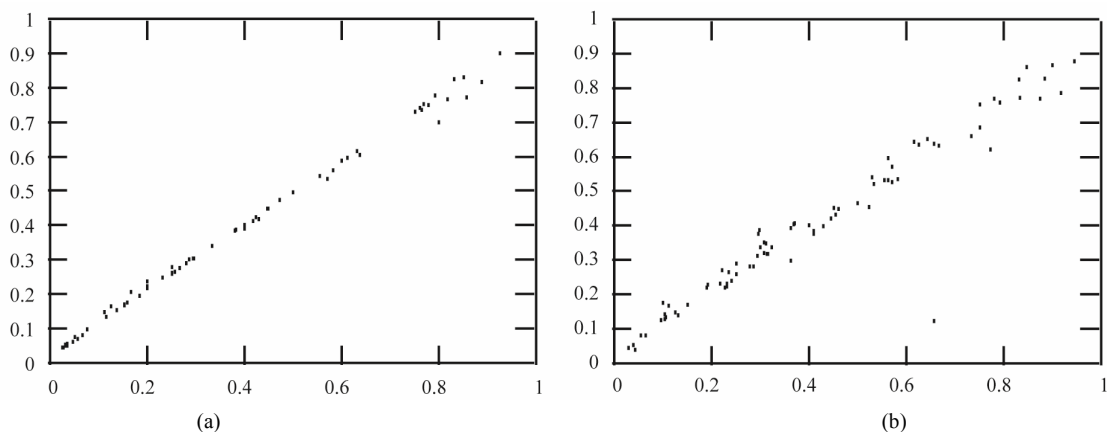


Figure 7. (a) Plot of \bar{p}_{31} versus \hat{p}_{31} ; (b) Plot of \tilde{p}_{31} versus \hat{p}_{31} .

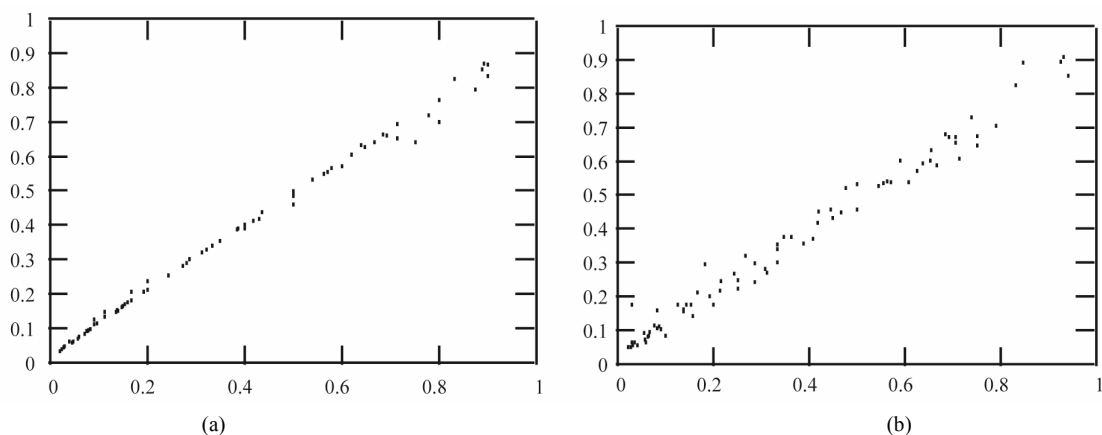


Figure 8. (a) Plot of \bar{p}_{32} versus \hat{p}_{32} ; (b) Plot of \tilde{p}_{32} versus \hat{p}_{32} .

simulation will be developed for different Bayesian models. The next step will be to proof analytically this property.

REFERENCES

- [1] P. Diaconis and D. Ylvisaker, "Conjugate Priors for Exponential Families," *The Annals of Statistics*, Vol. 7, No. 2, 1979, pp. 269-281.
- [2] T. C. Lee, G. G. Judge and A. Zellner, "Maximum Likelihood and Bayesian Estimation of Transition Probabilities," *JASA*, Vol. 63, No. 324, 1968, pp. 1162-1179.
- [3] S. Assoudou and B. Essebbar, "A Bayesian Model for Markov Chains via Jeffreys' Prior," Department of Mathematics and Computer Sciences, Faculté des Sciences of Rabat, Morocco, 2001.
- [4] C. Robert, "Méthode de Monte Carlo par Chanes de Markov," Economica, Paris, 1996.