CUTOFF AND HITTING TIMES BASED ON DIVERGENCE MEASURES

ALASSANE DIÉDHIOU[1] AND PAPA NGOM[2]

[1] *UFR Sciences et Technologie*
Université de Ziguinchor BP 523 Ziguinchor Sénégal
e-mail : asana@ucad.sn
[2] *LMA - Laboratoire de Mathématiques Appliauées*
Université Cheikh Anta Diop BP 5005 Dakar-Fann Sénégal
e-mail : pngom@ucad.sn

Abstract. The cutoff phenomenon in the simulated methods is widely investigated in recent years. A important question is to detect the stopping time after which, one can obtain the convergence to equilibrium of interesting Markov chain. We propose in this paper, a method for evaluating the cutoff instant using appropriate stopping times such as those found in Lachaud (2005). We give the conditions under which a Markov process has a cutoff time, in the sense of Reyni's divergence measure; which can be considered as a generalization of the well known Kullback distance. For illustration, we evaluate the effectiveness of our method on the Ornstein-Uhlenbeck process.

1. Introduction

We study in this paper the cutoff phenomenon for a sample of Ornstein-Uhlenbech processes and for the average process by using the Rényi divergence measure. There is a vast literature concerning cutoff phenomenon in the last twenty years, (see e.g. [2, 5]). As we will see in theorem (2.1) before a certain deterministic time called cutoff instant, the sample remains far from its equilibrium, then it converges exponentially fast. Asymptotically the cutoff is equivalent to the hitting time defined as the instant at which an empirical mean of the sample reaches its expected value for the first time see [5]. In [3] the cutoff is studied, by using the Kullback distance.
The novelty in his work, lies mainly in the fact that the Rényi divergence measure which we use here generalizes the Kullback distance. Let us denote Rényi divergence measure by d_R.
For two measures ν and μ dominated both by λ, and f and g the respective densities with respect to λ, the Rényi divergence measure is defined as:

$$d_R(\nu,\mu) = \Big(\frac{1}{\alpha(\alpha-1)} Log \int f^{\alpha} g^{1-\alpha} d\lambda\Big)^{\frac{1}{2}},\ \alpha \in \mathbb{R},\ \alpha \neq 0 \ and \ \alpha \neq 1.$$

It is well known that for the limiting cases $\alpha = 0$ or $\alpha = 1$, d_R tends to the Kullback-Leibler divergence see [1].
Our method is similar as that used by B. Lachaud and B. Ycart in [3]. In their work they use the Kullback distance, and they prove some asymptotic results of the behavior of the cutoff instant.

Key words and phrases. Cutoff, hitting time, divergence measure, test of goodness fit.

In our application of the theoretical results we use the test based on the Rényi measure.
The paper is organized as follows. Section 2 contains several preliminary results, in section 3 we study the asymptotic behavior of $R^{\alpha}(\widehat{\theta}_n, \theta_o)$ and in the last section we have the numerical results.

2. Preliminaries

We do the following assumptions. Let μ_t be the distribution of the process X at time t. We suppose that the process X_t converges exponentially fast to X_∞ with exponetial rate ρ. If we denote by μ_∞ the distribution of X_∞, we assume that there exist positive contants K and ρ such that

$$\lim_{t \longrightarrow +\infty} d_R(\mu_t, \mu_\infty) e^{\rho t} = K. \tag{2.1}$$

Let $X^{(n)} = (X_1, \cdots, X_n)$ be a $n-$sample of the process X. By the previous assumption each coordinate converges with exponential rate ρ to its asymptotic equilibrium.
Denoting by d_R^n the Rényi distance between the distribution of X^n at time t and the equilibrium distribution $\mu_\infty^{\otimes n}$, we have the

Theorem 2.1. [3] *Assume that* (2.1) *holds, and let u be a fixed real. Then*

$$\lim_{n \longrightarrow +\infty} d_R^n(u + \frac{logn}{2\rho}) = Ke^{-\rho u}.$$

This theorem gives the behavior of the sample with respect to the cutoff instant $\frac{logn}{2\rho}$ in function of the sign of the real u see [3].
Let us consider a test function f which is a mapping from the state space E into $\mathbb{R}$. And let us set

$$M_{x_0}^{(n)}(f(t)) = \frac{1}{n} \sum_{i=1}^{n} f(X_i)(t), \ t \geq 0.$$

We have

$$\mathbb{E}[M_{x_0}^{(n)}(f(t))] = \mathbb{E}[f(X)(t)], \ t \geq 0,$$

since the coordinates $X_i(t)$ are i.i.d. with distribution of X.
By the law of large numbers, one can prove that $M_{x_0}^{(n)}(f(t))$ converges a.s. to $\mathbb{E}[f(X)(t)]$ when n tends to infinity at all times. If the process X converges exponentially fast to X_∞ in the sense of total variation distance, the $\mathbb{E}[f(X)(t)]$ converges exponentially fast to $\mu_\infty f = \int f d\mu_\infty$ when t tends to infinity.
We now find out the first time the process $M_{x_0}^{(n)}(f(t))$ meets $\mu_\infty f$. So let us define the hitting time

$$T_{x_0}^n(f) = inf\{t \geq 0; M_{x_0}^{(n)}(f(t)) = \mu_\infty f\}.$$

In [5] it is shown that

$$\lim_{n \longrightarrow +\infty} \mathbb{P}(|\frac{T_{x_0}^n(f)}{\frac{Logn}{2\rho}} - 1| < \epsilon) = 1, \ \forall \epsilon > 0,$$

i.e. the hitting time is asymptotically equivalent to the cutoff $\frac{Logn}{2\rho}$.
We will apply this results to the Orsntein-Uhlenbeck process X which is the solution

of the following stochastic differential equation:

$$\begin{cases} dX_t = -\rho X_t dt + \sigma\sqrt{2\rho} dB_t, \\ X_0 = x_0 \end{cases} \tag{2.2}$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion, ρ and σ positive reals, and x_0 a real. This equation can be solved explicitly and we get

$$X_t = x_0 e^{-\rho t} + \sigma\sqrt{2\rho}\int_0^t e^{-\rho(t-s)} dB_s,\ t \geq 0.$$

The process $(X_t)_{t\geq 0}$ is Gaussian with expectation $x_0 e^{-\rho t}$ and variance $\sigma^2(1-e^{-2\rho t})$. So we can show that

$$X_t \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2),$$

when t tends to infinity.
For $\alpha \leq e^{2\rho t}$, and $\alpha \neq 0$ *and* 1 we have

$$d_R(\mu_t, \mu_\infty) = \Big(\frac{1}{\alpha(\alpha-1)}(\frac{\alpha}{4}Log(1-e^{-2\rho t}) - \frac{1}{4}Log(1-\alpha e^{-2\rho t})) + \frac{x_0^2 e^{-2\rho t}}{2\sigma^2}\Big)^{\frac{1}{2}},$$

hence we have

$$\lim_{t\longrightarrow+\infty} d_R(\mu_t, \mu_\infty)e^{\rho t} = \frac{|x_0|}{\sigma\sqrt{2}}.$$

For the OU process the theorem (2.1) holds, in fact

$$d_R^n(t) = \Big(\frac{1}{\alpha(\alpha-1)}(\frac{n\alpha}{4}Log(1-\frac{e^{-2\rho u}}{n}) - \frac{n}{4}Log(1-\frac{\alpha e^{-2\rho u}}{n})) + \frac{x_0^2 e^{-2\rho u}}{2\sigma^2}\Big)^{\frac{1}{2}},$$

then we have

$$\lim_{n\longrightarrow+\infty} d_R^n(u + \frac{Logn}{2\rho}) = \frac{|x_0|}{\sigma\sqrt{2}}e^{-\rho u}.$$

In the next sections, we shall discuss the practical consequences of these results for the detection of convergence.

3. Asymptotic behavior of $R^\alpha(\widehat{\theta}_n, \theta_o)$

For a measurable space (X, A) and for a family of probability distributions functions

$$F = \{F_\theta : \theta \in \Theta, \text{where } \Theta \text{ is an open set of } \mathbb{R}^d\},$$

dominated by a μ measure, $\sigma-$additive and $\sigma-$finite defined in a $\sigma-$algebra A, where,

$$f(x,\theta) = \frac{dF_\theta}{d\mu}$$

are density functions that fulfill the following regularity conditions :

(1) the $A = \{x \in X/f(x,\theta) > 0\}$ does not depend on θ and for all $x \in A, \theta \in \Theta$,

$$\frac{\partial f(x,\theta)}{\partial\theta_i},\quad \frac{\partial^2 f(x,\theta)}{\partial\theta_i\partial\theta_j},\quad \frac{\partial^3 f(x,\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k},\quad i,j,k = 1,\dots,d$$

exist and are finite.

(2) There exist real-valued functions $F(x)$ and $H(x)$ such that

$$|\frac{\partial f(x,\theta)}{\partial\theta_i}| < F(x),\quad |\frac{\partial^2 f(x,\theta)}{\partial\theta_i\partial\theta_j}| < F(x),\quad |\frac{\partial^3 f(x,\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}| < H(x),$$

where F is finitely integrate and $E[H(X)] < M$, with M independent of θ.

(3) The Fisher information matrix

$$I_F(\theta) = \left[E\left\{ \frac{\partial log f(X,\theta)}{\partial\theta_i} \frac{\partial log f(X,\theta)}{\partial\theta_j} \right\} \right]_{i,j=1,\dots,d}$$

is finite and positive definite.

Here, we choose an important measure of divergence given by Rényi which can be written in following form :

$$R^\alpha(\theta_1,\theta_2) = d_R^2(\theta_1,\theta_2) = \frac{1}{\alpha(\alpha-1)} \ln\left(\int_X f_{\theta_1}^\alpha(x) f_{\theta_2}^{1-\alpha}(x) d\mu(x) \right); \quad \alpha \neq 0,1$$

We suppose there exits a strongly consistent sequence $\widehat{\theta}_n$ of roots of likelihood equations such that :

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, I_F(\theta)^{-1}), \tag{3.1}$$

Assume that $(1)-(3)$ and (3.1) hold. Under $H_o : \theta \in \Theta_o \subset \Theta$, we present the asymptotic distribution of $R^\alpha(\widehat{\theta}_n, \theta_o)$.

Theorem 3.1. *Let the model and ϕ satisfy $(1)-(3)$ and (3.1) respectively. Let θ_o be the true parameter :*

(i) if $\theta \neq \theta_o$, we have,

$$\sqrt{n}\big[R^\alpha(\widehat{\theta}_n, \theta_o) - R^\alpha(\theta,\theta_o)\big] \to N[0, \sigma^2(\theta,\theta_o)],$$

where $\sigma^2(\theta,\theta_o) = AI_F(\theta)^{-1}A^t$ and $A = \nabla R^\alpha(\theta,\theta_o)$ with $\nabla = (\frac{\partial}{\partial\theta_1}, \dots, \frac{\partial}{\partial\theta_d})$.

(ii) if $\theta = \theta_o$, then we have,

$$2nR^\alpha(\widehat{\theta}_n, \theta_o) \to \sum_{i=1}^r \beta_i Z_i^2$$

where the Z_i are $N[0,1]$ and the β_i are of eigen value of $M\Sigma_\theta$ with $r = rank(\Sigma_\theta M \Sigma_\theta)$ with $M = \left(\frac{\partial^2 R^\alpha(\theta,\theta_o)}{\partial\theta_i \partial\theta_j} \right)_{i,j=1,\dots,d}$

Proof. we suppose that R^α has continuous second partial derivatives verifing regularity conditions (1)-(3):

(i) If $\theta \neq \theta_o$.
A first order Taylor expansion gives

$$R^\alpha(\widehat{\theta}_n, \theta_o) = R^\alpha(\theta,\theta_o) + A(\widehat{\theta}_n - \theta)^t + o(\|\widehat{\theta}_n - \theta\|).$$

As

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, I_F(\theta)^{-1}),$$

it is clear that the random variables,
$\sqrt{n}\big[R^\alpha(\widehat{\theta}_n, \theta_o) - R^\alpha(\theta,\theta_o)\big]$ and $A\sqrt{n}(\hat{\theta}_n - \theta)^t$ have the same asymptotic distribution, because

$$\sqrt{n}\ o\ (\|\hat{\theta}_n - \theta\|) = o_p(1).$$

(ii) If $\theta = \theta_o$
A Taylor series expansion of $R^\alpha(\widehat{\theta}_n, \theta)$ around $\widehat{\theta}_n$ yields

$$R^{\alpha}(\widehat{\theta}_n,\theta_o)=R^{\alpha}(\theta,\theta_o)+\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\frac{\partial^2 R^{\alpha}(\widehat{\theta}_n,\theta_o)}{\partial\theta_i\partial\theta_j}(\widehat{\theta}_{ni}-\theta_i)(\widehat{\theta}_{nj}-\theta_j)+o(||\widehat{\theta}_n-\theta||^2)$$

Therefore $R^{\alpha}(\theta,\theta_o)=0$:

$$2nR^{\alpha}(\widehat{\theta}_n,\theta)=n(\widehat{\theta}_n-\theta)M(\theta)(\widehat{\theta}_n-\theta)^t$$

has the same distribution as

$$M=\left(\frac{\partial^2 R^{\alpha}(\theta,\theta_o)}{\partial\theta_i\partial\theta_j}\right)_{i,j=1,\dots,d}$$

In another part,

$$\sqrt{n}(\widehat{\theta}_n-\theta)\to N[0,\Sigma_\theta]$$

where

$$\Sigma_\theta=diag(\theta)-\theta\theta^t$$

We have:

$$n(\widehat{\theta}_n-\theta)M(\theta)(\widehat{\theta}_n-\theta)^t\to\sum_{i=1}^{r}\beta_i Z_i^2$$

where the Z_i are $N[0,1]$ and the β_i are of eigen value of $M\Sigma_\theta$ with $r=rank(\Sigma_\theta M\Sigma_\theta)$ □

Remark 3.1. *On the basis has theorem 3.1, we get an approximation to the power function $P(\theta)=P(R^{\alpha}(\widehat{\theta}_n,\theta)>\chi^2_d)$ in the following way :*

$$P(\theta)=1-F_{N(0,1)}\left[\frac{\sqrt{n}}{\sigma(\theta,\theta_o)}\left(\frac{\chi^2_d,\alpha}{2n}-R^{\alpha}(\theta,\theta_o)\right)\right]$$

This result can be used to obtain an approximation to the power of $R^{\alpha}(\widehat{\theta}_n,\theta)$ test in the same way as in Theorem 3.1.

The previous results giving the asymptotic distribution of the $R^{\alpha}(\widehat{\theta}_n,\theta)$ divergence statistic in random sampling can be used to test statistical hypotheses.

Let $LX^n(t)$ the distribution of X^n at time t. We want to test :

$$\begin{cases} H_o: & LX^n(t)=\mu_\infty \\ & \text{vs} \\ H_1: & LX^n(t)\neq\mu_\infty. \end{cases}$$

The following section illustrates this fact.

4. A numerical study

For the detection of the instant of convergence of a stochastic process to its asymptotic distribution, we consider the following method, for which, the implementation requires the following computations :

(1) choose an integer n and build a $n-$sample of the process;
(2) run the $n-$sample, and compute the mean process at each time;
(3) stop when the mean process reaches for the first time the expectation of the asymptotic distribution.

To illustrate the method proposed here, we consider an example. The sampled process is the Ornstein-Uhlenbeck diffusion described in the previous section, with following parameters : $\rho = 1$; $\sigma = 1$; and $x_o = 10$.

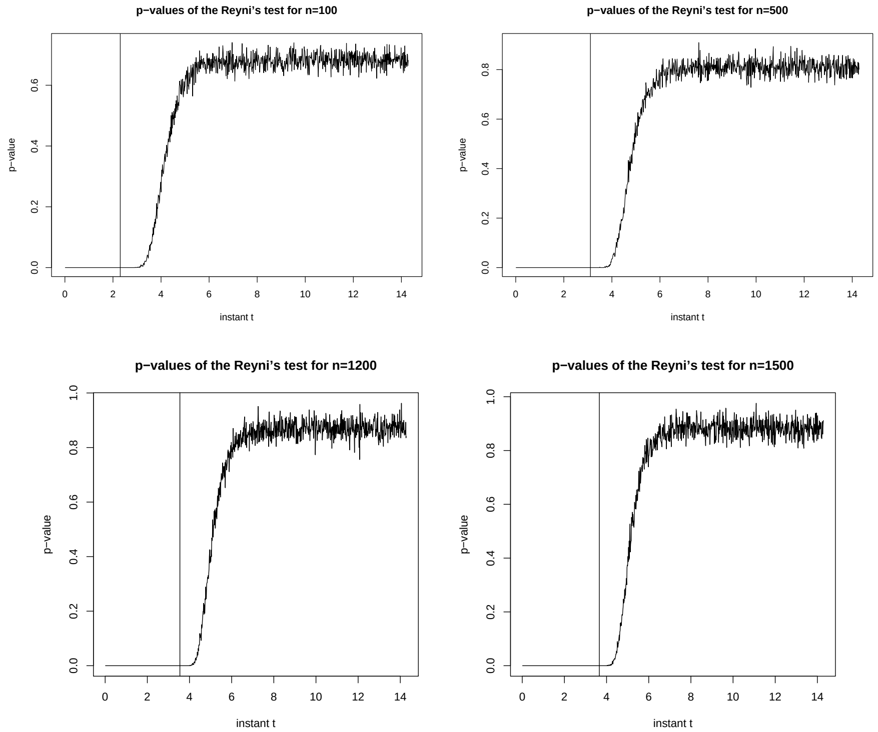

Figure 1 : Average of the p-value of the Reyni's test for different values of n.

in our numerical study, we consider various sets of experiements in which the data are generated from the Ornstein-Uhlenbek process. The sample size varies from 100 to 1200, and for each sample size the number of replications is 1,000. We consider the following methodology based on the study of evolution of the process and its convergence.
Our procedure is :

$*$ we fix a sequence of regularity spaced $t_1, t_2, \ldots, t_M$ with $t_i = i/70$ and $i = 1, \ldots, M = 1,000$.

$*$ the experience consists in simulating a $n-$sample of Ornstein-Uhlenbeck process $(X_1, X_2, \ldots, X_n))$ with the following parameters $\rho = 1, \sigma = 1$ and $x_0 = 10$.
At each instant t_i, we test if our sample process $(X_1(t_i), X_2(t_i), \ldots, X_n(t_i))$ is a Gaussian sample with $N(0, \sigma^2)$.

We use the Reyni's test statistic of goodness-of-fit. The p-value of the test is computed for each instant t_i. Note that, on the sample experiment, the first instant where the mean of the n-sample reaches zero is computed ; it is an observation of r.v T_{x_0}.
The cutoff instant is materialized by the vertical line in the figures. In order to

check whether the equilibrium is indeed obtained, the experiment is run after the hitting time. After 1,000 replications of the experiment, for each t_i, we have a 1,000-sample of p-values. The average of the 1,000 p-values associated to each t_i, are presented in following figures.
The results of our four sets of experiments are presented in figures 1-4.
The first figure reports the case when the p-values are drawn for $n = 100$. Here the vertical line corresponds to the cutoff instant $\dfrac{\log n}{2\rho} \simeq 2.302$.
For figures 2, 3 and 4, the p-values are generated from $n = 500$, 1200 and 1500, with the cutoff instant 2.303, 3.107, 3.545 and 3.657 respectively.

We observed that the averaged p-value of the Reyni's test is null before the cutoff instant on, and becomes positive after this instant. Furthermore, the null hypothesis is systematically rejected before the cutoff instant.
We wait until approximatively twice the cutoff instant, and we observe that the stationary regime is established when the average p-value stabilizes around 80%.

5. Conclusion

In summary, similarly to the classical notions, we have introduced the Reyni distance, for detecting the convergence of a stochastic process to its asymptotic distribution. We have investigated the fundamental property of steep convergence to equilibrium of Ornstein-Uhlenbeck (OU) process.
One of things that we have tried to establish here, is that there is a close link between the cutoff phenomenon and the detection of convergence, in the case of Ornstein-Uhlenbek process.
We restrict our study to the case of OU, but the simulation results in section 4, indicate that the theorem 2.1 does extend to more general process.
Our study, illustrated by numerical results, has shown that the method proposed here, detect very well the the convergence of the process to its asymptotic distribution, in the sense of Reyni's distance. It appears clearly that the convergence occurs after the hitting time of the asymptotic expectation by an empirical mean.

References

[1] D. Morales, L. Pardo *New smooth test statistics of goodness of fit for categorized composite null hypotheses*, Sociedad de Estadística e Investigación Operativa, Test(2000) Vol.**9**, pp.173-190.
[2] B. Lachaud, *Cutoff and hitting times for a sample of Ornstein-Uhlenbeck processes and its evarage*, J. Appl. Probab. **42**(4)(2005) 1069-1080.
[3] B. Lachaud, B. Ycart, *Convergence times for parallel Markov chains*
[4] M. Salicrù, D. Morales, M. L. Menendez and L. Pardo, *On the applications of divergence type measures in testing statistical hypotheses* Journal of multivariate analysis, 51, 372-391 (1994).
[5] B. Ycart, *Stopping tests for Markov chain monte-Carlo methods*, methodol. Comput.Appl. Probab.**2(1)** (2000) 23-36.