



# TEST ON REGENERATIVE PROCESS THROUGH $\Phi$ -DIVERGENCE

Jacques Touboul

## ► To cite this version:

| Jacques Touboul. TEST ON REGENERATIVE PROCESS THROUGH  $\Phi$ -DIVERGENCE. 2008. <hal-00312006>

**HAL Id: hal-00312006**

**<https://hal.science/hal-00312006v1>**

Preprint submitted on 24 Aug 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# TEST ON REGENERATIVE PROCESS THROUGH $\Phi$ -DIVERGENCE

**Jacques TOUBOUL**

*Jack\_touboul@hotmail.com*

*Laboratoire de Statistique Théorique et Appliquée*

*Université Pierre et Marie Curie*

*175, rue du Chevaleret, 75013 Paris, France*

---

## Abstract

The purpose of this article is to expose a new sufficient condition to reject the hypothesis saying that a process is a regenerative process. This condition is based on  $\Phi$ -divergence and on the fact that its associated renewal process follows a parametric model.

*Key words:* Regenerative Process; Renewal Process; minimum  $\Phi$ -divergence; Point Process.  
*2000 MSC:* 60K20 62H15 62H12 60G51

---

## Introduction

We consider  $(X_t)_{t \in T}$  a random process - where  $T = \mathbb{R}$  or  $T = \mathbb{N}$  - and we suppose that  $(X_t)_{t \in T}$  is a regenerative process. In this paper, our goal will be to write a test which verifies, in real time, if this hypothesis is acceptable through  $\Phi$ -divergences.

Let us then consider the renewal process associated with  $(X_t)_{t \in T}$  that we will name  $(\tau_n)_{n \in \mathbb{N}}$ . This process is a random walk on  $\mathbb{R}^+$ . We will consequently suppose here that the law of  $(\tau_n - \tau_{n-1})$  belongs to the stochastic model  $\mathcal{M} = \{P_\lambda; \lambda \in \Lambda\}$  - where  $\Lambda$  is an open of  $\mathbb{R}$ . Let  $\lambda_0$  be the true parameter.

In a first part, we provide an estimator  $\hat{\lambda}_n$  of  $\lambda_0$  and we prove the uniform convergence of  $f_{\hat{\lambda}_n}$  to  $f_{\lambda_0}$  where, more generally,  $f_\lambda$  is the density of the probability  $P_\lambda \in \mathcal{M}$ . In a second part, we carry out a test on  $\lambda_0$  and conclude our theory. Finally, we perform a simulation.

All reminders and proofs can be found in annex.

## 1. Convergences

Based on works [1] and [2], we derive estimators of  $\lambda_0$ . Then, after introducing certain notations, we will produce almost sure uniform convergences of these expressions.

### 1.1 Writing the estimators

We consider an identifiable parametric model  $\{P_\lambda; \lambda \in \Lambda\}$  defined on some measurable space

$(\mathcal{X}, \mathcal{B})$  and  $\Lambda$  is an open of  $\mathbb{R}$ . We assume for all  $\lambda$  in  $\Lambda$ ,  $P_\lambda$  has a density  $f_\lambda$  with respect to some dominating  $\sigma$ -finite measure. From an i.i.d. sample  $T_1, T_2, \dots, T_n$  with distribution  $P_{\lambda_0}$ , we aim at estimating  $\lambda_0$ , the true value of the parameter.

Now, let us introduce the concept of  $\Phi$ -divergence. Let  $\varphi$  be a strictly convex function defined by  $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ , and such that  $\varphi(1) = 0$ . We define a  $\Phi$ -divergence of  $P$  from  $Q$  - where  $P$  and  $Q$  are two probability distributions over a space  $\Omega$  such that  $Q$  is absolutely continuous with respect to  $P$  - by  $\Phi(Q, P) = \int \varphi(\frac{dQ}{dP}) dP$ . Moreover, let  $\varphi^*$  be a function defined by,  $\forall t \in \mathbb{R}$ ,  $\varphi^*(t) = t\varphi'^{-1}(t) - \varphi(\varphi'^{-1}(t))$ , where  $\varphi'$  is the derivate function of  $\varphi$ ,  $\varphi'^{-1}$  the reciprocal function of  $\varphi'$  and let  $\mathcal{F}$  be the class of function defined by  $\mathcal{F} = \{x \mapsto \varphi'(\frac{f_\alpha}{f_\lambda}); \lambda \in \mathbb{R}^+\}$ , then [1] and [2] show that the estimator of  $\Phi(P_\alpha, P_{\lambda_0})$  - that we will call  $\Phi(\alpha, \lambda_0)$ - is :

$$\hat{\Phi}_n(\alpha, \lambda) = \sup_{\lambda \in \Lambda} \left\{ \int \varphi'(\frac{f_\alpha}{f_\lambda}) dP_\alpha - \int \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda})) dP_n \right\},$$

where  $P_n$  is the empirical measure of  $(T_n)$  and thus the minimum  $\Phi$ -divergence estimate of  $\lambda_0$  is :

$$\hat{\lambda}_n = \arg \inf_{\alpha \in \Lambda} \hat{\Phi}_n(\alpha, \lambda).$$

## 1.2 Convergence studies

Let us consider  $\Lambda_\alpha = \{\lambda \in \Lambda \mid \int \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda})) dP_{\lambda_0} < \infty\}$ ,  
 $M(\lambda, \alpha, x) = \int \varphi'(\frac{f_\alpha}{f_\lambda}) dP_\alpha - \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda}))$ ,  $P_n M(\lambda, \alpha) = \int \varphi'(\frac{f_\alpha}{f_\lambda}) dP_\alpha - \int \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda})) dP_n$ ,  
 $PM(\lambda, \alpha) = \int \varphi'(\frac{f_\alpha}{f_\lambda}) dP_\alpha - \int \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda})) dP$ ,  $\hat{c}_n(\alpha) = \arg \sup_{\lambda \in \Lambda} P_n M(\lambda, \alpha)$ ,  
 $\tilde{c}_n(\alpha) = \arg \sup_{\lambda \in \Lambda_\alpha} P_n M(\lambda, \alpha)$ ,  $\hat{\gamma}_n = \arg \inf_{\alpha \in \Lambda} \sup_{\lambda \in \Lambda} P_n M(\lambda, \alpha)$  and  
 $\tilde{\gamma}_n = \arg \inf_{\alpha \in \Lambda} \sup_{\lambda \in \Lambda_\alpha} P_n M(\lambda, \alpha)$ .

We remark that  $\hat{\lambda}_n$  is a  $M$ -estimator for  $\lambda_0$  and its rate of convergence is consequently in  $O_P(m^{-1/2})$ . However, *Van der Vaart*, in chapter V of his work [8], thoroughly studies  $M$ -estimators and formulates hypotheses that we will use here in our context and for all set  $\lambda_0$ :

- (H1) :  $\sup_{\alpha \in \Lambda; \lambda \in \Lambda_\alpha} |P_n M(\lambda, \alpha) - PM(\lambda, \alpha)| \rightarrow 0$  a.s. (respectively in probability)
- (H2) : For all  $\varepsilon > 0$ , there is  $\eta > 0$ , such that for all  $\lambda \in \Lambda_\alpha$  verifying  $\|\lambda - \lambda_0\| \geq \varepsilon$ , we have  $PM(\lambda, \alpha) - \eta > PM(\lambda_0, \alpha)$ , with  $\alpha \in \Lambda$ .
- (H3) :  $\exists Z < 0$ ,  $n_0 > 0$  such that  $(n \geq n_0 \Rightarrow \sup_{\alpha \in \Lambda} \sup_{\lambda \in \{\Lambda_\alpha\}^c} P_n M(\lambda_0, \alpha) < Z)$
- (H4) : There is a neighbourhood of  $\lambda_0$ ,  $V$ , and a positive function  $H$ , such that, for all  $\lambda \in V$  we have  $|M(\lambda, \lambda_0, x)| \leq H(x)$  ( $P$ -p.s.) with  $PH < \infty$ ,
- (H5) : There is a neighbourhood  $V$  of  $\lambda_0$ , such that for all  $\varepsilon$ , there is a  $\eta$  such that for all  $\lambda \in V$  and  $\alpha \in \Lambda$ , verifying  $\|\alpha - \lambda_0\| \geq \varepsilon$ , we have  $PM(\lambda, \lambda_0) < PM(\lambda, \alpha) - \eta$ .

According to Broniatowski, we can thus say that:

**Proposition 1. : Assuming conditions (H1) to (H5) hold, we have**

- (1)  $\sup_{\alpha \in \Lambda} \|\hat{c}_n(\alpha) - \lambda_0\|$  tends to 0 a.s. (respectively in probability)
- (2)  $\hat{\gamma}_n$  tends to  $\lambda_0$  a.s. (respectively in probability).

Finally, if  $n$  is the number of vectors of the sample, we then have

**Theorem 1. :** We have almost everywhere and even uniformly almost everywhere, the following convergence:  $f_{\lambda_n} \rightarrow f_{\lambda_0}$ , when  $n \rightarrow \infty$ .

## 2. Test

Taking note of section 1, let us consider  $T_n^\Phi$  the function defined by  $T_n^\Phi(\alpha, \lambda_0) = \frac{2n}{\varphi''(1)} \hat{\Phi}_n(\alpha, \lambda_0)$ , where  $\hat{\Phi}_n(\alpha, \lambda) = \sup_{\alpha \in \Lambda} \{ \int \varphi'(\frac{f_\lambda}{f_\alpha}) dP_\lambda - \int \varphi^*(\frac{f_\lambda}{f_\alpha}) dP_n \}$ . The articles [7] and [9] show that this function converges towards a  $\chi^2$  random variable if  $\alpha = \lambda_0$ . Hence, since  $\varphi$  is a positive function, we can write a rupture detection test for any  $(T_n)$  parameter, i.e.  $H0 : \lambda = \lambda_0$  versus  $H1 : \lambda \neq \lambda_0$ , through the function  $T_n^\Phi(\alpha, \lambda_0)$ , i.e. by the critical region  $\mathcal{R}^\Phi = \{ \frac{2n}{\varphi''(1)} \hat{\Phi}_n(\alpha_0, \lambda_0) > q_{1-\varepsilon} \}$ , where  $q_{1-\varepsilon}$  is the quantile, of level  $1 - \varepsilon$ , of a  $\chi^2$  distribution and where, under  $(H0)$ ,  $\alpha_0$  is the unique element such that  $\Phi(\alpha_0, \lambda_0) = 0$  according to proposition 2 (see page 5).

Finally, if  $(H0)$  is not acceptable, the hypothesis stating that  $(X_t)$  is a regenerative process is also not acceptable, since the independance of variables  $(Z_t)$  - which is defined page 4 within the definition of a regenerative process - no longer holds.

## 3. Simulation

In this section, we take a real point process (see definition page 4) as the renewal process associated with  $(X_t)_{t \in T}$ .

First, we simulate a point process such that its parameter  $\lambda_0$  is equal to 1 and we will estimate  $\lambda_0$ . Second, we will randomly change the parameter and we will observe when the rupture can be detected. We obtain

Estimate of the $\lambda_0$	theoretical value :	1
	estimate :	0.980055
	P-Value :	0.719
When the parameter changes :	theoretical value :	20
	detection :	after 88 random variables generated
	estimate :	19.7074
	P-Value :	0.9

### Critics of the simulation :

We note that as the approximations accumulate and according to the power of the calculators used, we might obtain results above or below the value of the thresholds of the different tests. Moreover, in the case where  $\lambda_0$  is unknown, we will never be sure to have reached the minimum of the  $\Phi$ -divergence: we have indeed used the simulated annealing method to solve our optimisation problem, and therefore it is only when the number of random jumps tends in theory towards infinity that the probability to get the minimum tends to 1. We note finally that no theory on the optimal number of jumps to implement does exist, as this number depends on the specificities of each particular problem.

### Conclusion :

The present article demonstrates that our  $\Phi$ -divergence method constitutes a good sufficient condition to reject the hypothesis according to a process is a regenerative process. Indeed, the convergence results and simulations we carried out, convincingly fulfilled our expectations.

## Annex A - Reminders

In this section, we briefly recall the concepts that we need in this article :

### A.1. Regenerative Process, Associated Renewal Process And Point Process :

Let  $T$  be the set equal to  $\mathbb{R}$  or  $\mathbb{N}$ . Let  $(X_t)_{t \in T}$  be a random process such that  $X_t \in E$ , where  $E$  is a topological set, and such that the applications  $t \rightarrow X_t$  are right-continuous. We have

**Definition 1.** The process  $(X_t)_{t \in T}$  is said to be generative, if there exists an increasing sequence  $(\tau_n)_{n \in \mathbb{N}}$  of random variables such that  $\tau_n \in T$  and for any measurable function  $f : E \rightarrow \mathbb{R}^+$ , the random variables

$$Z_n = \int_{\tau_n}^{\tau_{n+1}} f(X_s) ds \text{ if } T = \mathbb{R}^+, Z_n = \sum_{k=\tau_n}^{\tau_{n+1}-1} f(X_k) \text{ if } T = \mathbb{N},$$

are independent and identically-distributed.

We remark that if we take  $f = 1$ , we can say that the random variables  $(\tau_{n+1} - \tau_n)_n$  are positive, independent and identically-distributed. Hence, the sequence  $(\tau_n)_n$  is a random walk on  $\mathbb{R}^+$ . We will call  $(\tau_n)_n$  the renewal process associated with its regenerative process  $(X_t)_{t \in T}$ .

Moreover, we have

**Definition 2.** We define a point process on  $\mathbb{R}^+$  as the sequence of random variable  $(T_i)$  such that  $0 \leq T_1 < T_2 < \dots$ ,  $T_i \in \mathbb{R}^+$  and  $T_n \rightarrow \infty$  a.s..

and finally

**Theorem 2.** Let  $(T_i)$  be a point process. Thus, the random variables  $T_1, T_2 - T_1, T_3 - T_2, \dots$  are mutually independent and have the same exponential distribution.

### A.2. $\Phi$ -Divergences

Let  $\varphi$  be a strictly convex function defined by  $\varphi : \overline{\mathbb{R}^+} \rightarrow \overline{\mathbb{R}^+}$ , and such that  $\varphi(1) = 0$ .

**Definition 3.** We define  $\Phi$ -divergence of  $P$  from  $Q$  - where  $P$  and  $Q$  are two probability distributions over a space  $\Omega$  such that  $Q$  is absolutely continuous with respect to  $P$  - by

$$\Phi(Q, P) = \int \varphi\left(\frac{dQ}{dP}\right) dP.$$

It will be noted that this expression also holds if  $P$  and  $Q$  are both dominated by the same probability.

The most used distances (Kullback, Hellinger or  $\chi^2$ ) belong to the Cressie-Read family (see [4], [3] and the book [6]). They are defined by a specific  $\varphi$ . Indeed,

- with the relative entropy, we associate  $\varphi(x) = x \ln(x) - x + 1$
- with the Hellinger distance, we associate  $\varphi(x) = 2(\sqrt{x} - 1)^2$
- with the  $\chi^2$  distance, we associate  $\varphi(x) = \frac{1}{2}(x - 1)^2$
- more generally, with power divergences, we associate  $\varphi(x) = \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)}$ , where  $\gamma \in \mathbb{R} \setminus (0, 1)$
- and, finally, with the  $L^1$  norm, which is also a divergence, we associate  $\varphi(x) = |x - 1|$ .

Finally, we have

**Proposition 2.** *A fundamental property of  $\Phi$ -divergences is the fact that there is a unique case of nullity. We have  $\Phi(P, Q) = 0 \Leftrightarrow P = Q$ .*

## Annex B - PROOFS

### Proof of proposition 1 :

Given that  $X_n \rightarrow X$  (a.s.) if  $\forall \varepsilon > 0$ ,  $P(\limsup\{|X_n - X| > \varepsilon\}) = 0$ , we prove proposition 1:

**Proof :** Since  $\tilde{c}_n(\alpha) = \arg \sup_{\lambda \in \Lambda_n} P_n M(\lambda, \alpha)$ , we have  $P_n M(\tilde{c}_n(\alpha), \alpha) \geq P_n M(\lambda_0, \alpha)$ . And through condition (H1), we get  $P_n M(\tilde{c}_n(\alpha), \alpha) \geq P_n M(\lambda_0, \alpha) \geq PM(\lambda_0, \alpha) - o_{P_n}(1)$ , where  $o_{P_n}(1)$  does not depend on  $\alpha$ . Thus, we get:

$$\begin{aligned} PM(\lambda_0, \alpha) - P_n M(\tilde{c}_n(\alpha), \alpha) &\leq P_n M(\tilde{c}_n(\alpha), \alpha) - PM(\tilde{c}_n(\alpha), \alpha) + o_{P_n}(1) \\ &\leq \sup_{\alpha \in \Lambda; \lambda \in \Lambda_n} |P_n M(\lambda, \alpha) - PM(\lambda, \alpha)| \rightarrow 0 \text{ a.s. } (*). \end{aligned}$$

Let  $\varepsilon > 0$  be such that  $\sup_{\alpha \in \Lambda} \|\tilde{c}_n(\alpha) - \lambda_0\| > \varepsilon$ . We notice that if such  $\varepsilon$  had failed to exist, the result would be obvious. Therefore, for this  $\varepsilon$ , there is  $a_n \in \Lambda$  such that  $\|\tilde{c}_n(a_n) - \lambda_0\| > \varepsilon$ , which implies thanks to (H2) that there exists a  $\eta$  such that  $PM(\tilde{c}_n(a_n), a_n) - PM(\lambda_0, a_n) > \eta$ . Thus, we can write :

$$P(\sup_{a \in \mathbb{R}^d} \|\tilde{c}_n(\alpha) - \lambda_0\| > \varepsilon) \leq P(PM(\tilde{c}_n(a_n), a_n) - PM(\lambda_0, a_n) > \eta) \rightarrow 0 \text{ by } (*).$$

Moreover, (H1) and (H3) imply that  $\hat{c}_n(\alpha) = \tilde{c}_n(\alpha)$  for all  $\alpha \in \Lambda$  and for  $n$  big enough. This results in  $\sup_{\alpha \in \Lambda} \|\hat{c}_n(\alpha) - \lambda_0\| \rightarrow 0$  a.s., which concludes our demonstration of the first part of the proposition.

For the second part, we remark that (H1) and (H3) also imply that  $\hat{\gamma}_n = \tilde{\gamma}_n$  for all  $\alpha \in \Lambda$ . This explains why it is sufficient to demonstrate the result for  $\tilde{\gamma}_n$  only.

Based on the first part of the demonstration and on condition (H4), we can write:

$$P_n M(\tilde{c}_n(\tilde{\gamma}_n), \tilde{\gamma}_n) \geq P_n M(\tilde{c}_n(\lambda_0), \lambda_0) \geq PM(\tilde{c}_n(\tilde{\gamma}_n), \lambda_0) - o_{P_n}(1),$$

which implies:

$$\begin{aligned} PM(\tilde{c}_n(\tilde{\gamma}_n), \lambda_0) - PM(\tilde{c}_n(\tilde{\gamma}_n), \tilde{\gamma}_n) &\leq P_n M(\tilde{c}_n(\tilde{\gamma}_n), \tilde{\gamma}_n) - PM(\tilde{c}_n(\tilde{\gamma}_n), \tilde{\gamma}_n) + o_{P_n}(1) \\ &\leq \sup_{a \in \Lambda; b \in \Lambda_n} |P_n M(\lambda, \alpha) - PM(\lambda, \alpha)| \rightarrow 0 \text{ a.s. } (**). \end{aligned}$$

Based on the first part of this demonstration and on (H5), we infer the existence of  $\eta$  such that :  $P(\|\tilde{\gamma}_n - \lambda_0\| \geq \varepsilon) \leq P(PM(\tilde{c}_n(\tilde{\gamma}_n), \lambda_0) - PM(\tilde{c}_n(\tilde{\gamma}_n), \tilde{\gamma}_n)) \rightarrow 0 \text{ a.s. by } (**),$  which concludes our demonstration.  $\square$

**Proof of theorem 1 :** Let  $F_\lambda$  be the cumulative distribution function of a probability which belongs to our parametric model  $\mathcal{M}$  and let  $\psi_\lambda$  be a complex function defined by  $\psi_\lambda(u, v) = F_\lambda(\text{Re}(u + iv)) + iF_\lambda(\text{Re}(v + iu))$ , for all  $u$  and  $v$  in  $\mathbb{R}$ .

First, according to proposal (9.1) page 216 of the book [5], "Any defined and continuously differentiable, in an open set  $D \subset \mathbb{C}$ , complex function is analytical in  $D$ " we can therefore say that  $\psi_\lambda(u, v)$  is an analytic function, because  $x \mapsto f_\lambda(x)$  is a continuous function.

Given the corollary of Dini's second theorem - according to which "A sequence of cumulative distribution functions which simply converges on  $\mathbb{R}$  towards a continuous cumulative distribution function  $F$  on  $\mathbb{R}$ , uniformly converges towards  $F$  on  $\mathbb{R}$ " - we deduct that, for all sequence  $(\lambda_n)$  converging towards  $\lambda$ ,  $\psi_{\lambda_n}$  uniformly converges toward  $\psi_\lambda$ .

Finally, the Weierstrass theorem, (see proposal (10.1) page 220 of [5]), states that "Let  $(f_n)$  be a sequence of analytic function in an open set  $D \subset \mathbb{C}$ , and let us suppose that for every closed disc  $\Delta$  included in  $D$ , the sequence  $(f_n(z))$  uniformly converges in  $\Delta$  toward a limit  $f(z)$ . Hence  $f$  is an analytic function in  $D$ , and for all  $k \geq 1$ , the sequence of derivative functions  $(f_n^{(k)}(z))$  uniformly converges in  $\Delta$  towards  $(f^{(k)}(z))$ ." Applying the above reasoning to  $\psi_\lambda$ , we derive for

$k = 1$ , that all sequence  $\psi'_{\lambda,n}$  uniformly converge towards  $\psi'_\lambda$ , for all  $\lambda_n$  tending to  $\lambda$ . We can therefore conclude.

## References

- [1] Broniatowski M., *Estimation of the Kullback-Leibler divergence*, *Math. Methods Statist.*, 12(4):391–409(2004), 2003.
- [2] Broniatowski M. and Amor Keziou, *Dual representation of  $\phi$ -divergences and applications*, *C. R. Math. Acad. Sci. Paris*, 336(10):857–862, 2003.
- [3] Csiszár, I. *On topology properties of  $f$ -divergences*. *Studia Sci. Math. Hungar.* 2 1967 329–339.
- [4] Cressie, Noel; Read, Timothy R. C. *Multinomial goodness-of-fit tests*. *J. Roy. Statist. Soc. Ser. B* 46 (1984), no. 3, 440–464.
- [5] Jean Dieudonné, *Calcul infinitésimal*. Hermann.
- [6] Liese Friedrich and Vajda Igor, *Convex statistical distances*, volume 95 of *Teubner-Texte zur Mathematik [Teubner Texts in Mathematics]*. BSB B. G. Teubner Verlagsgesellschaft, 1987, with German, French and Russian summaries.
- [7] Pardo, Leandro, *Statistical inference based on divergence measures*, volume 185 of *Statistics: Textbooks and Monographs*. Chapman & Hall/CRC, Boca Raton, FL, 2006.
- [8] van der Vaart A. W., *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, 1998.
- [9] Zografos, K. and Ferentinos, K. and Papaioannou, T.,  *$\phi$ -divergence statistics: sampling properties and multinomial goodness of fit and divergence tests*, *Comm. Statist. Theory Methods*, 19(5):1785–1802(1990), MR1075502.