



HAL
open science

CONTROL OF THE AUTOMATIC OPENING OF A SERVER THROUGH Φ -DIVERGENCE

Jacques Touboul

► **To cite this version:**

Jacques Touboul. CONTROL OF THE AUTOMATIC OPENING OF A SERVER THROUGH Φ -DIVERGENCE. 2008. hal-00315875

HAL Id: hal-00315875

<https://hal.science/hal-00315875>

Preprint submitted on 1 Sep 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.

CONTROL OF THE AUTOMATIC OPENING OF A SERVER THROUGH Φ -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 provide a new starting control test of a new server in a queue. This control is based on Φ -divergence.

Key words: Queueing theory; minimum Φ -divergence; Point Process.

2000 MSC: 60K25 62H15 62H12 60G51

Introduction

Let us consider a queue. Let us assume that its arrival process distribution abides by the stochastic model $\mathcal{M} = \{P_\lambda; \lambda \in \Lambda\}$ - where Λ is an open of \mathbb{R} . Let λ_0 be the true parameter. We suppose also that we have one server and that the service process distribution is independent from P_{λ_0} . Now, let us assume that another server must start. The purpose of the present article will be to produce, in real time, a Φ -divergence-based test to verify that this start has indeed occurred.

In the first part of this article, we will provide an estimator $\hat{\lambda}_n$ of λ_0 and we will prove there is uniform convergence of $f_{\hat{\lambda}_n}$ to f_{λ_0} where, more generally, f_λ is the density of the probability $P_\lambda \in \mathcal{M}$. Then, we will carry out a test on λ_0 and conclude our theory. Finally, we will perform a simulation.

All reminders and proofs can be found in annex.

1. Convergences

Based on works [1] and [2], we derive estimators of λ_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 Λ is an open of \mathbb{R} . We assume for all λ in Λ , P_λ has a density f_λ with respect to some

dominating σ -finite measure. From an i.i.d. sample T_1, T_2, \dots, T_n with distribution P_{λ_0} , we aim at estimating λ_0 , the true value of the parameter.

Now, let us introduce the concept of Φ -divergence. Let φ be a strictly convex function defined by $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, and such that $\varphi(1) = 0$. We define a Φ -divergence of P from Q - where P and Q are two probability distributions over a space Ω such that Q is absolutely continuous with respect to P - by

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

Moreover, let φ^* be a function defined by, $\forall t \in \mathbb{R}$, $\varphi^*(t) = t\varphi^{-1}(t) - \varphi(\varphi^{-1}(t))$, where φ' is the derivate function of φ , φ'^{-1} the reciprocal function of φ' 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 :

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

where P_n is the empirical measure of (T_n) and thus the minimum Φ -divergence estimate of λ_0 is

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

1.2 Convergence studies

Let us consider $\Lambda_\alpha = \{\lambda \in \Lambda \mid \int \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda})) dP_{\lambda_0} < \infty\}$,
 $M(\lambda, \alpha, x) = \int \varphi'\left(\frac{f_\alpha}{f_\lambda}\right) dP_\alpha - \varphi^*\left(\varphi'\left(\frac{f_\alpha}{f_\lambda}\right)\right)$, $P_n M(\lambda, \alpha) = \int \varphi'\left(\frac{f_\alpha}{f_\lambda}\right) dP_\alpha - \int \varphi^*\left(\varphi'\left(\frac{f_\alpha}{f_\lambda}\right)\right) dP_n$,
 $PM(\lambda, \alpha) = \int \varphi'\left(\frac{f_\alpha}{f_\lambda}\right) dP_\alpha - \int \varphi^*\left(\varphi'\left(\frac{f_\alpha}{f_\lambda}\right)\right) dP$, $\hat{c}_n(\alpha) = \arg \sup_{\lambda \in \Lambda} P_n M(\lambda, \alpha)$,
 $\tilde{c}_n(\alpha) = \arg \sup_{\lambda \in \Lambda_\alpha} P_n M(\lambda, \alpha)$, $\hat{\gamma}_n = \arg \inf_{\alpha \in \Lambda} \sup_{\lambda \in \Lambda} P_n M(\lambda, \alpha)$ 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 λ_0 and its rate of convergence is consequently in $O_P(m^{-1/2})$. However, *Van der Vaart*, in chapter V of his work [10], thoroughly studies M -estimators and formulates hypotheses that we will use here in our context and for all set λ_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 λ_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 λ_0 , such that for all ε , there is a η 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 λ_0 a.s. (respectively in probability).

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

Theorem 1. :

We have uniformly almost everywhere the following convergence: $f_{\hat{\lambda}_n} \rightarrow f_{\lambda_0}$, when $n \rightarrow \infty$.

2. Test

Taking note of section 1, let us consider T_n^Φ 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 [9] and [11] show that this function converges towards a χ^2 random variable if $\alpha = \lambda_0$. Hence, since φ 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 χ^2 distribution and where, under $(H0)$, α_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, we can conclude that the second server has indeed started.

3. Simulation

In this section, we take a real point process (see definition page 4) as the arrival process.

First, we simulate a point process such that its parameter λ_0 is equal to 1000 and we will estimate λ_0 . Second, we will change the parameter and we will observe when the rupture can be detected. We obtain

Estimate of the λ_0	theoretical value :	1000
	estimate :	995.5
	P-Value :	0.989
When the parameter changes :	theoretical value :	2000
	detection :	after 40 random variables generated
	estimate :	2006
	P-Value :	0.998

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 λ_0 is unknown, we will never be sure to have reached the minimum of the Φ -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 Φ -divergence method constitutes a good way of controlling concerning the starting off a new server in a queue. 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. Process with Independent and Stationary Increments (P.I.S.I.), Point Process and Queueing Theory:

Let us introduce a generalization of the Bernoulli process defined on \mathbb{N} onto \mathbb{R}^+ . Let us then consider the notion of process with independent and stationary increments.

Definition 1. Let $\{X_t; t \in \mathbb{R}^+\}$ be a process with real values. This process is called a process with independent and stationary increments (P.I.S.I.), if

- 1/ The applications $t \rightarrow X_t$ are right-continuous on \mathbb{R}^+ ,
- 2/ For all $s, t \leq 0$, $X_t - X_s$ is independent from $\sigma(X_r; r \geq s)$,
- 3/ For all $s, t \leq 0$, $X_{t+s} - X_s$ has the same law as $X_t - X_0$ and
- 4/ We have $X_0 = 0$.

Then, 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.. The countable process - associated with this point process - is the sequence (N_t) such that $N_t = \sum_{k=1}^{\infty} \mathbf{1}_{[0,t]}(T_k)$ $t \in \mathbb{R}^+$.

and

Theorem 2. Let (T_i) be a point process such that its associated countable process is a P.I.S.I.. Thus, the random variables $T_1, T_2 - T_1, T_3 - T_2, \dots$ are mutually independent and have the same exponential distribution with $\lambda > 0$ parameters. Moreover, for all $t \leq 0$, N_t is a Poisson random variable with parameter λt .

Finally, [7], [8] and [12] provide us with an introduction to the queueing theory. Let us just recall what is needed in our here.

Definition 3. The three most important characteristics of a queueing system can be found below:

1/ Arrival Process $(\tau_n)_n$:

Refers to the sequence of random variables describing customer arrivals in the system. It is interesting to note that, if there are no information on $(\tau_n)_n$, we can assume that the process $(Tn)_n$, where $Tn = \tau_{n+1} - \tau_n$, is a point process.

2/ Service Process :

Refers to the sequence of random variables describing time devoted to customer service in the system. We often suppose that this distribution is independent from the distribution of $(Tn)_n$.

3/ Number of Servers :

Number of servers available to service the customers. This number is not necessary finite.

A.2. Φ -Divergences

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

Definition 4. We define Φ -divergence of P from Q - where P and Q are two probability distributions over a space Ω 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 χ^2) belong to the Cressie-Read family (see [4], [3] and the book [6]). They are defined by a specific φ . 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 χ^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 Φ -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 α . 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 ε had failed to exist, the result would be obvious. Therefore, for this ε , 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 η such that $PM(\tilde{c}_n(a_n), a_n) - PM(\lambda_0, a_n) > \eta$. Thus, we can write :

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

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:

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

Based on the first part of this demonstration and on (H5), we infer the existence of η 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_λ be the cumulative distribution function of a probability which belongs to our parametric model \mathcal{M} and let ψ_λ 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 (λ_n) converging towards λ , ψ_{λ_n} uniformly converges toward ψ_λ .

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 Δ included in D , the sequence $(f_n(z))$ uniformly converges in Δ 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 Δ towards $(f^{(k)}(z))$." Applying the above reasoning to ψ_λ , we derive for $k = 1$, that all sequence ψ'_{λ_n} uniformly converge towards ψ'_λ , for all λ_n tending to λ . We can therefore conclude.

References

- [1] Broniatowski M., *Estimation of the Kullback-Leibler divergence*, *Math. Methods Stat.*,12(4):391–409(2004), 2003.
- [2] Broniatowski M. and Amor Keziou, *Dual representation of ϕ -divergences and applications*, *C. R. Math. Acad. Sci. Paris*,336(10):857–862, 2003.
- [3] Csizsár, 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] Gross, Donald; Carl M. Harris, *Fundamentals of Queueing Theory*.1998. Wiley.
- [8] Lazowska, Edward D.; John Zahorjan, G. Scott Graham, Kenneth C. Sevcik. *Quantitative System Performance: Computer System Analysis Using Queueing Network Models*. 1984. Prentice-Hall, Inc.
- [9] Pardo, Leandro, *Statistical inference based on divergence measures*, volume 185 of *Statistics: Textbooks and Monographs*. Chapman & Hall/CRC, Boca Raton, FL, 2006.
- [10] van der Vaart A. W., *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge, 1998.
- [11] Zografos, K. and Ferentinos, K. and Papaioannou, T., *ϕ -divergence statistics: sampling properties and multinomial goodness of fit and divergence tests*, *Comm. Statist. Theory Methods*,19(5):1785–1802(1990), MR1075502.
- [12] Zukerman, Moshe. *Introduction to Queueing Theory and Stochastic Teletraffic Models*. 2008. <http://www.ee.unimelb.edu.au/staff/mzu/classnotes.pdf>