TEST OF RUPTURE DETECTION THROUGH Φ-DIVERGENCE

Jacques Touboul

STATISTICS

TEST OF RUPTURE DETECTION THROUGH Φ-DIVERGENCE

Jacques Touboul

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 test of rupture detection for point process based on Φ-divergences.

Résumé

Test de détection de rupture par Φ-diverence. Le but de cet article est de présenter un nouveau test de rupture pour les processus ponctuels fondés sur les Φ-divergences.

Version Française Abrégée

Considérons (T_n), un processus ponctuel de paramètre λ_0. Puisque toute Φ-divergence s'écrit de manière unique sous la forme $\Phi(P,Q) = \int \varphi(\frac{dP}{dQ})dQ$ - où P et Q sont deux probabilités telles que P est absolument continue par rapport à Q et où φ est une fonction réelle, strictement convexe et telle que $\varphi(1) = 0$ - nous établissons tout d'abord un estimateur de λ_0. En effet, soit Φ une divergence, en posant $\Lambda = \mathbb{R}^+$, P_n la mesure empirique de (T_n) et, pour tout réel t, $\varphi^*(t) = t\varphi'^{-1}(t) - \varphi(\varphi'^{-1}(t))$, où φ' est la dérivée de φ et où φ'^{-1} est la fonction réciproque de φ', nous montrons que
$\hat{\lambda}_n = arg\inf_{\alpha\in\Lambda}\sup_{\lambda\in\Lambda}\{\int\ \varphi'(\frac{f_\lambda}{f_\alpha})\ dP_\lambda\ -\ \int\ \varphi^*(\varphi'(\frac{f_\lambda}{f_\alpha}))\ dP_n\}$ converge presque sûrement vers λ_0. Puis, si f_λ est la densité d'une loi exponentielle de paramètre λ, nous montrons que $f_{\hat{\lambda}_n}$ converge uniformément vers f_{λ_0} et que, pour tout réel x, $|f_{\hat{\lambda}_n}(x) - f_{\lambda_0}(x)| = O_P(n^{-1/2})$. Considérons maintenant la fonction $T_n^\Phi(\alpha,\lambda_0)$ définie par $T_n^\Phi(\alpha,\lambda_0) = \frac{2n}{\varphi''(1)}\hat{\Phi}_n(\alpha,\lambda_0)$, où $\hat{\Phi}_n(\alpha,\lambda) = \sup_{\alpha\in\Lambda}\{\int\ \varphi'(\frac{f_\lambda}{f_\alpha})\ dP_\lambda\ -\ \int\ \varphi^*(\varphi'(\frac{f_\lambda}{f_\alpha}))\ dP_n\}$. Alors, Leandro Pardo (voir [Pardo, Leandro, 2006]) et Zografos (1990) (voir [Zografos, K. and Ferentinos, K. and Papaioannou, T., 1990]) ont montré que $T_n^\Phi(\alpha,\lambda_0)$ convergeait en loi vers une variable aléatoire de loi χ^2. Et c'est ainsi que nous construisons un test basé sur la région critique $\mathcal{R}^\Phi = \{\frac{2n}{\varphi''(1)}\hat{\Phi}_n(\alpha_0,\lambda_0) > q_{1-\varepsilon}\}$, où $q_{1-\varepsilon}$ est le quantile de niveau $1-\varepsilon$ d'une loi χ^2.

Email address: Jack_touboul@hotmail.com (Jacques Touboul).

Preprint submitted to a journal *August 25, 2008*

Enfin, nous effectuons des simulations et nous obtenons une détection de rupture d'intensité au bout de 62 variables aléatoires générées, ce qui nous montre l'intérêt de ce test dans beaucoup de situations comme par exemple l'arrivée de voyageurs en gare ou le traitement en temps réel de la distribution d'électricité.

Introduction

We will define point process as randomly distributed points from a time and/or space standpoint.
In this paper, we will consider a real point process and our goal will be to study the modifications of λ_0 through Φ-divergences. All reminders and proofs can be found in annex.

1. Convergences

Let (T_n) be a point process and let λ_0 be its parameter.
Based on the work of Broniatowski in [Broniatowski, 2003] and [Broniatowski and Keziou, 2003], 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 $\sigma-$finite measure. From an i.i.d. sample X_1, X_2,...,X_n with distribution P_{λ_0}, we aim at estimating λ_0, the true value of the parameter. We first need to focus on the selection of the most appropriate model. Indeed, on the one hand, for any $n > 0$, $T_n - T_{n-1}$ is an exponential random variable and on the other hand, $T_n = \Sigma(T_i - T_{i-1})$, with $T_0 = 0$, has a gamma distribution. Since we want to study the variation of λ_0, we decide to adopt the exponential model, ie we take $X_0 = T_0$, $X_1 = T_1 - T_0$, $X_n = T_n - T_{n-1}$ and $f_{\lambda_0}(x) = \lambda_0 exp\{-\lambda_0 x\}$.
Now, let us introduce the concept of $\Phi-$divergence. Let φ be a strictly convex function defined by $\varphi : \overline{\mathbb{R}^+} \to \overline{\mathbb{R}^+}$, and such as $\varphi(1) = 0$. We define a $\Phi-$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 $\Phi(Q, P) = \int \varphi(\frac{dQ}{dP})dP$. 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 Broniatowski in [Broniatowski, 2003] and [Broniatowski and Keziou, 2003] shows that the estimator of $\Phi(P_\alpha, P_{\lambda_0})$ - that we will call $\Phi(\alpha, \lambda_0)$- is :
$\hat{\Phi}_n(\alpha, \lambda) = \sup_{\lambda\in\Lambda}\{\int\ \varphi'(\frac{f_\alpha}{f_\lambda})\ dP_\alpha\ -\ \int\ \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda}))\ dP_n\}$,
where P_n is the empirical of (X_n) measure 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 $\Lambda_\alpha = \{\lambda \in \Lambda \mid\ \int \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda}))\ dP_{\lambda_0} < \infty\}$, $M(\lambda, \alpha, x) = \int \varphi'(\frac{f_\alpha}{f_\lambda})dP_\alpha -\ \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda}))$, $P_nM(\lambda, \alpha) = \int \varphi'(\frac{f_\alpha}{f_\lambda})dP_\alpha -\ \int \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda}))dP_n$, $PM(\lambda, \alpha) = \int \varphi'(\frac{f_\alpha}{f_\lambda})dP_\alpha -\ \int \varphi^*(\varphi'(\frac{f_\alpha}{f_\lambda}))dP$,
$\hat{c}_n(\alpha) =\ arg \sup_{\lambda\in\Lambda}\ P_nM(\lambda, \alpha)$, $\tilde{c}_n(\alpha) =\ arg \sup_{\lambda\in\Lambda_\alpha}\ P_nM(\lambda, \alpha)$, $\hat{\gamma}_n =\ arg \inf_{\alpha\in\Lambda}\ \sup_{\lambda\in\Lambda}\ P_nM(\lambda, \alpha)$ and $\tilde{\gamma}_n =\ arg \inf_{\alpha\in\Lambda}\ \sup_{\lambda\in\Lambda_\alpha}\ P_nM(\lambda, \alpha)$.
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 [van der Vaart, 1998], thoroughly studies M-estimators

and formulates hypotheses that we will use here in our context and for all set λ_0:

$$(H1)\ :\ \sup_{\alpha\in\Lambda;\ \lambda\in\Lambda_\alpha} |P_nM(\lambda,\alpha)-PM(\lambda,\alpha)| \to 0\ a.s.\ (respectively\ in\ probability)$$

$$(H2)\ :\ For\ all\ \varepsilon>0,\ there\ is\ \eta>0,\ such\ that\ for\ all\ \lambda\in\Lambda_\alpha\ verifying\ \|\lambda-\lambda_0\|\geq\varepsilon,$$
$$we\ have\ PM(\lambda,\alpha)-\eta>PM(\lambda_0,\alpha),\ with\ \alpha\in\Lambda.$$

$$(H3)\ :\ \exists Z<0,\ n_0>0\ such\ that\ (n\geq n_0\ \Rightarrow\ \sup_{\alpha\in\Lambda}\sup_{\lambda\in\{\Lambda_\alpha\}^c} P_nM(\lambda_0,\alpha)<Z)$$

$$(H4)\ :\ There\ is\ a\ neighbourhood\ of\ \lambda_0,\ V,\ and\ a\ positive\ function\ H,\ such\ that,$$
$$for\ all\ \lambda\in V\ we\ have\ |M(\lambda,\lambda_0,x)|\leq H(x)\ (P-p.s.)\ with\ PH<\infty,$$

$$(H5)\ :\ There\ is\ a\ neighbourhood\ V\ of\ \lambda_0,\ such\ that\ for\ all\ \varepsilon,\ there\ is\ a\ \eta\ such\ that\ for$$
$$all\ \lambda\in V\ and\ \alpha\in\Lambda,\ verifying\ \|\alpha-\lambda_0\|\geq\varepsilon,\ we\ have\ PM(\lambda,\lambda_0)<PM(\lambda,\alpha)-\eta.$$

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 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 almost everywhere and even uniformly almost everywhere, the following convergence: $f_{\hat{\lambda}_n}\to f_{\lambda_0}$, when $n\to\infty$.

2. Rate of convergence

Theorem 2 : For all real x, we have $|f_{\hat{\lambda}_n}(x)-f_{\lambda_0}(x)|=O_P(n^{-1/2})$.

3. Test

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 $\hat{\Phi}_n(\alpha,\lambda)=\sup_{\alpha\in\Lambda}\{\int\ \varphi'(\frac{f_\lambda}{f_\alpha})\,dP_\lambda\ -\ \int\ \varphi^*(\frac{f_\lambda}{f_\alpha})\,dP_n\}$, then many mathematicians, such as Zografos (1990) (see [Zografos, K. and Ferentinos, K. and Papaioannou, T., 1990]) and Leandro Pardo (see [Pardo, Leandro, 2006]), have shown 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 new rupture detection test for any (T_n) intensity, 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).

4. Simulation

First, we simulate a point process such that its parameter is equal to $\lambda_0=1$ and we will estimate λ_0. Second, we will randomly change the parameter and we will observe when the rupture can be detected.

We obtain

Estimate of the λ_0	theoritical value :	1
	estimate :	0.977367
	P-Value :	0.704
When the parameter changes :	theoritical value :	32.4
	detection :	after 62 random variables generated
	estimate :	32.4528
	P-Value :	0.999

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 test of rupture detection. Indeed, the convergence results and simulations we carried out, convincingly fulfilled our expectations. We believe this test could be of great use for many industries such as transportation, for instance to deal with incoming passenger flows or for utilities to distribute electricity in real time.

Annex A - Reminders
In this section, we briefly recall the concepts that we will need below :

A.1. *Process with Independent and Stationary Increments* (*P.I.S.I.*) *and point process :*

Let us introduce a generalization of the Bernouilli 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 $t \to X_t$ 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 $0 \leq T_1 < T_2 < ....$, $T_i \in \mathbb{R}^+$ and $T_n \to \infty$ a.s.. The countable process - associated with this point process - is the sequence (N_t) such that $N_t = \Sigma_{k=1}^{\infty} \mathbf{1}_{[0,t]}(T_k)$ $t \in \mathbb{R}^+$.*
and finally

Theorem 3 *Let (T_i) be a point process such that its associated countable process is a P.I.S.I.. Thus, the random variables $T_1, T_2 - T_1, T_3 - T_2, ...$ 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.*

A.2. Φ-*Divergences*

Let φ be a strictly convex function defined by $\varphi : \overline{\mathbb{R}^+} \to \overline{\mathbb{R}^+}$, 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 Ω such that Q is absolutely continuous with respect to P - by $\Phi(Q, P) = \int \varphi(\frac{dQ}{dP})dP$. 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 Csiszar 1967, Cressie - Read 1984 and the book [Friedrich and Igor, 1987]). They are defined by a specific φ. Indeed,

- with the relative entropy, we associate $\varphi(x) = xln(x) - x + 1$
- 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 $\Phi-$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 $X_n \to X$ (a.s.) if $\forall \varepsilon > 0$, $P(\lim sup\{|X_n - X| > \varepsilon\}) = 0$, we prove proposition 1:

Proof : Since $\tilde{c}_n(\alpha) = arg \sup_{\lambda \in \Lambda_\alpha} P_n M(\lambda, \alpha)$, 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:

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

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 :

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

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 $\sup_{\alpha \in \Lambda} \|\hat{c}_n(\alpha) - \lambda_0\| \to 0$ *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:

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

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

Proof of theorem 1 : Let F_a be the cumulative distribution function of a exponential random variable such that its parameter is a and ψ_a is a complex function defined by
$\psi_a(u,v) = F_a(Re(u+iv)) + iF_a(Re(v+iu))$, for all u and v in $\mathbb{R}$.
First, according to proposal (9.1) page 216 of the book "Calcul Infinitésimal" of Jean Dieudonne, *"Any defined and continuously differentiable, in an open set $D \subset \mathbb{C}$, complex function is analytical in D."* We can therefore say that $\psi_a(u,v)$ is an analytic function, because $x \mapsto f_a(x) = ae^{-ax}$ 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 (a_n) converging towards a, ψ_{a_n} uniformly converges toward ψ_a. Finally, the Weierstrass theorem, (see proposal (10.1) page 220 of the "Calcul infinitésimal" book of Jean Dieudonné), 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_a, we derive for $k = 1$, that all sequence $\psi'_{a,n}$ uniformly converge towards ψ'_a, for all a_n tending to a. We can therefore conclude.

Proof of theorem 2 : For all x, we have

$$\begin{aligned}
|f_{\hat{\lambda}_n}(x) - f_{\lambda_0}(x)| &= |\hat{\lambda}_n e^{-\hat{\lambda}_n x} - \lambda_0 e^{-\lambda_0 x}| = |(\hat{\lambda}_n - \lambda_0)e^{-\hat{\lambda}_n x} + \lambda_0(e^{-\hat{\lambda}_n x} - e^{-\lambda_0 x})| \\
&= |(\hat{\lambda}_n - \lambda_0)e^{-\hat{\lambda}_n x} + \lambda_0 e^{-\lambda_0 x}(e^{-(\hat{\lambda}_n - \lambda_0)x} - 1)| \\
&= |(\hat{\lambda}_n - \lambda_0)e^{-\hat{\lambda}_n x} + \lambda_0 e^{-\lambda_0 x}(1 - (\hat{\lambda}_n - \lambda_0)x + o(x) - 1)| \\
&= |(\hat{\lambda}_n - \lambda_0)e^{-\hat{\lambda}_n x} + \lambda_0 x e^{-\lambda_0 x}(-(\hat{\lambda}_n - \lambda_0) + o(1))| \\
&\leq |\hat{\lambda}_n - \lambda_0| e^{-\hat{\lambda}_n x} + \lambda_0 x e^{-\lambda_0 x} |(\hat{\lambda}_n - \lambda_0)| + |o(1))| \\
&\leq O_P(n^{-1/2})O_P(1) + O_P(1)O_P(n^{-1/2}) + |o(1)| = O_P(n^{-1/2}),
\end{aligned}$$

since $\hat{\lambda}_n$ is a M-estimator for λ_0 and its rate of convergence is consequently in $O_P(n^{-1/2})$. Hence, we get the result.

References

[Broniatowski, 2003] Broniatowski M., *Estimation of the Kullback-Leibler divergence, Math. Methods Statist.,12(4):391–409(2004), 2003.*

[Broniatowski and Keziou, 2003] Broniatowski M. and Amor Keziou, *Dual representation of ϕ-divergences and applications, C. R. Math. Acad. Sci. Paris,336(10):857–862, 2003.*

[Friedrich and Igor, 1987] 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.*

[Pardo, Leandro, 2006] Pardo, Leandro, *Statistical inference based on divergence measures, volume 185 of Statistics: Textbooks and Monographs. Chapman & Hall/CRC, Boca Raton, FL, 2006.*

[van der Vaart, 1998] van der Vaart A. W., *Asymptotic statistics, volume 3 of Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 1998.*

[Zografos, K. and Ferentinos, K. and Papaioannou, T., 1990] 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.*