Un theorème de Donsker et de Glivenko-Cantelli pour une classe de processus généralisant le processus empirique

Davit VARRON [1]

[1] *Laboratoire de mathématique de Besançon,*
Université de Franche-Comté,
16 route de Gray,
25000 Besançon
davit.varron@univ-fcomte.fr

Résumé. Les mesures de probabilité discrètes aléatoires admettent de façon évidente une représentation de la forme $\sum_{i\geq 1} p_i\delta_{X_i}$, où les poids $(p_i)_{i\geq 1}$ et les $(X_i)_{i\geq 1}$ sont aléatoires. En statistique Bayesienne non paramétrique, on considére parfois de tels objets, vérifiant en plus la propriété suivante : conditionnellement aux p_i, les X_i sont independantes et identiquement distribuées. C'est par exemple le cas du fameux processus de Dirichlet. Ce type particulier de mesures de probabilité aléatoires constitue une extension intéressante de la mesure empirique. En effet, beaucoup d'outils sont transposables à cet objet plus général, au prix de quelques efforts techniques. Nous montrons, sous des conditions générales classiques de la théorie de processus empirique, un théorème de Glivenko-Cantelli et un théorème de de Donsker pour des suites de mesures de probabilité aléatoires de ce type. En corollaire de ces résultats, dans le cadre de la statistique bayésienne non paramétrique, nous prouvons un théorème de Bernstein-Von Mises pour les lois à postériori du processus du Dirichlet.

Mots-clés. Processus empiriques, Statistique Bayésienne non paramétrique.

Abstract. Every random probability measure can be represented as $\sum_{i\geq 1} p_i\delta_{X_i}$, where both the weights $(p_i)_{i\geq 1}$ and the atoms $(X_i)_{i\geq 1}$ are random. In Bayesian nonparamtric statistics, several priors also satisfy the following property : conditionnally to $(p_i)_{i\geq 1}$, the sequence $(X_i)_{i\geq 1}$ is independent and identically distributed. This is the case for the well known Dirichlet process. This particular kind of random probability measures is itself a very natural extension of the empirical measure, in the sense that several techniques in emprical process theory can be extendend to that wider class of objects. Under classical assumptions from the empirical processes theory, we establish a Glivenko-Cantelli and a Donsker theorem for sequences of such random probability measures. As a corollary, we establish a Berstein-Von Mises theorem for posterior distributions of the Dirichlet processe prior.

Keywords. Empirical processes, Bayesian Nonparametrics.

1 Introduction

Les propriétés asymptotiques de la mesure empirique ont été étudiés depuis de nombreuses décennies (voir par ex. Van der Vaart and Wellner (1996)). Parmi ces resultat, l'un des plus utilisés en statistique non paramétrique est que toute classe de fonctions admettant un nombre d'entropie uniforme $\sqrt{\log}$-intégrable, et admettant une envelope de carré intégrable est une classe de Donsker (Van der Vaart and Wellner, 1996, Theorem 2.5.2, p. 127). Dans cet exposé nous allons montrer que ces conditions structurelles sont suffisantes pour établir un théorème de Donsker et de Glivenko-Cantelli pour une classe plus générale de mesures aléatoires, que nous allons tout de suite décrire. Pour $\mathbf{p} = (p_i)_{i\geq 1} \in \mathbb{R}^{\mathbb{N}}$ et $r > 0$, nous noterons

$$|| \mathbf{p} ||_r := \Big(\sum_{i\geq 1} | p_i |^r \Big)^{1/r}, \tag{1.1}$$

qui peut éventuellement être infinie. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé complété et soit

$$\mathbb{S} := \Big\{ \mathbf{p} = (p_i)_{i\geq 1} \in [0, +\infty[^{\mathbb{N}},\ || \mathbf{p} ||_1 < \infty \Big\}, \tag{1.2}$$

le cône des suites positives sommable, équipé de sa tribu borélienne. Considérons une suite de collections de mesures de probabilité $\big(\mathbf{P}_{n,\mathbf{p}}\big)_{n\geq 1, \mathbf{p}\in\mathbb{S}}$ sur un espace mesurable $\big(\mathfrak{X}, \mathcal{X}\big)$. Pour n fixé, on définit la variable aléatoire $\boldsymbol{\beta}_n = (\beta_{i,n})_{i\geq 1}$ de $(\Omega, \mathcal{A}, \mathbb{P})$ vers $(\mathbb{S}, Borliens)$ et une suite $\mathbf{Y}_n = \big(Y_{i,n}\big)_{i\geq 1}$ de variables aléatoires de $(\Omega, \mathcal{A}, \mathbb{P})$ vers $\big(\mathfrak{X}, \mathcal{X}\big)$, pour laquelle, pour $\boldsymbol{\beta}_n$-presque tout $\mathbf{P}$, la loi conditionnelle sachant $\boldsymbol{\beta}_n = \mathbf{p}$ est la loi $\mathbf{P}_{n,\mathbf{p}}^{\otimes\mathbb{N}}$ (nous supposons également que ces lois conditionnelles sont définies au moyen d'un noyau de transition de Markov). Définissons enfin l'application suivante sur Ω.

$$Pr_n := \sum_{i\geq 1} \beta_{i,n} \delta_{Y_{i,n}}, \tag{1.3}$$

qui a toutes les propriétés d'une mesure de probabilité aléatoire. Il est intéressant d'obtenir des résultats asymptotiques (lorsque $n \to \infty$) pour Pr_n, pour au moins deux raisons. La première est que Pr_n est une généralisation de la mesure empirique. La seconde raison est que ce type d'objet est rencontré en statistique Bayésienne non paramétrique. L'exemple le plus typique peut être décrit de la façon suivante. Notons $\big(\mathfrak{M}, \mathcal{M}\big)$ pour l'ensemble des mesures de probabilité sur $\big(\mathfrak{X}, \mathcal{X}\big)$, équipée de la tribu borélienne $\mathcal{M}$ engendrée par la topologie faible. Étant donnée une loi de probabilité $\overline{\alpha}$ sur $\big(\mathfrak{X}, \mathcal{X}\big)$ et un paramètre de concentration $M > 0$, la distribution sur $\big(\mathfrak{M}, \mathcal{M}\big)$, notée $DP(M, \overline{\alpha})$, d'un processus de Dirichlet, admet la représentation suivant due à Sethuraman (1994) :

$$DP(M, \overline{\alpha}) =_d \sum_{i\geq 1} \beta_i \delta_{Y_i}. \tag{1.4}$$

Ici les Y_i sont i.i.d. de loi $\overline{\alpha}$ et $\beta_i := U_i \prod_{j=1}^{i-1}(1-U_j)$, où $(U_i)_{i\geq 1}$ est i.i.d de loi $Beta(1, M)$, indépendant des $(Y_i)_{i\geq 1}$. Il est également connu (Ferguson, 1973, Section 3, Theorem 1) que, si Pr a pour loi $DP(M, \overline{\alpha})$ et si $(X_1, \ldots, X_n)$ a pour loi $\mathbf{P}^{\otimes n}$ sachant $Pr = \mathbf{P}$ (pour Pr presque tout $\mathbf{P}$), alors une version de la loi à postériori de Pr sachant $(X_1, \ldots, X_n)$ est l'application

$$(x_1, \ldots, x_n) \to DP\Big(M+n, \theta_n\overline{\alpha} + (1-\theta_n)\overline{\alpha}_{(x_1,\ldots,x_n)}\Big), \tag{1.5}$$

où $\theta_n := M/(M+n)$ et $\overline{\alpha}_{(x_1,\ldots,x_n)} := n^{-1}\sum_{i\leq n}\delta_{x_i}$. Une classe plus générale de mesures de probabilité aléatoires est celle des mesures complètement aléatoires normalisées, sans atomes déterministes (Hjort et al. (2010, p. 84-85)).
Soit maintenant $\mathcal{F}$ une classe de fonctions réelles boréliennes, partout finies sur $(\mathfrak{X}, \mathcal{X})$. Définir proprement l'objet aléatoire $(\int f dPr_n)_{f\in\mathcal{F}}$ n'est pas immediat. Nous commencerons par supposer que $\mathcal{F}$ est séparable pour la topologie de la convergence point par point (voir par exemple Van der Vaart and Wellner (1996, p. 110)), et considérerons un séparant $\mathcal{F}_0$. Nous noterons $\mathcal{B}(\mathcal{F})$ l'espace des fonctions réelles bornées sur $\mathcal{F}$ qui sont continues par rapport à la topologie engendrée par les applications $\big\{f \to f(x),\ x \in \mathcal{X}\big\}$. Nous noterons aussi la norme sup usuelle sur $\mathcal{B}(\mathcal{F})$

$$||\ \psi\ ||_{\mathcal{F}} := \sup_{f\in\mathcal{F}} |\ \psi(f)\ |\ .$$

Nous noterons F l'enveloppe minimale de $\mathcal{F}$, c'est à dire $F:\ x \to \sup\big\{f(x),\ f\in\mathcal{F}_0\big\} \vee 1$. Pour $\epsilon > 0$, $r > 0$ et une loi de probabilité Q, nous noterons $N\big(\epsilon, \mathcal{F}, ||\cdot||_{Q,r}\ \big)$ pour le nombre minimal de boules de $L^r(Q)$ de rayon ϵ necessaire pour couvrir $\mathcal{F}$. Nous noterons également $\ell^\infty(\mathcal{F}) \supset \mathcal{B}(\mathcal{F})$ pour l'espace des fonctions rélles bornées sur $\mathcal{F}$. Pour $r > 0$, on définit l'espace $\mathcal{E}_{\mathcal{F},r}$ de la façon suivante : une application $\Psi : \Omega \to \mathcal{B}(\mathcal{F})$ appartient à $\mathcal{E}_{\mathcal{F},r}$ si et seulement si $\Psi(f)$ est Borel mesurable sur $(\Omega, \mathcal{A}, \mathbb{P})$ pour tout $f \in \mathcal{F}$, et si $|||\ \Psi\ |||_{\mathcal{F},r}^r := \mathbb{E}\big(\ ||\ \Psi\ ||_{\mathcal{F}}^r\ \big) < \infty$. Sous l'hypothèse que $\mathbb{E}\big(F(Y_{1,n})\big) < \infty$ pour tout n, il est possible de définir

$$G_n(\cdot):\ f \to \sum_{i\geq 1}\beta_{i,n}\Big(f(Y_{i,n}) - \mathbb{E}\big(f(Y_{i,n})\mid \boldsymbol{\beta}_n\big)\Big)$$

comme la limite, lorsque $k\to\infty$, des sommes tronquées

$$G_n^k(\cdot):\ f \to \sum_{i=1}^{k}\beta_{i,n}\Big(f(Y_{i,n}) - \mathbb{E}\big(f(Y_{i,n})\mid \boldsymbol{\beta}_n\big)\Big)$$

dans l'espace de Banach $\big(\mathcal{E}_{\mathcal{F},1}, |||\cdot|||_{\mathcal{F}}\ \big)$. Notre premier résultat est le suivant

Théorème 1 *Supposons que* $\| \boldsymbol{\beta}_n \|_1 = 1$ *presque sûrement pour tout* n, *et que* $\| \boldsymbol{\beta}_n \|_2 \to 0$ *en probabilité. Supposons que :*

$$\lim_{M\to\infty} \overline{\lim_{n\to\infty}} \, \mathbb{E}\Big(F(Y_{1,n}) \mathbb{1}_{\{F(Y_{1,n})\geq M\}} \Big) = 0. \tag{1.6}$$

Supposons également que, pour tous $\epsilon > 0$ *et* $M > 0$, *on a , lorsque* $n \to \infty$ *:*

$$\log\Big(N\big(\epsilon, \mathcal{F}_M, \| \cdot \|_{\overline{P}(\boldsymbol{\beta}_n, \mathbf{Y}_n),1} \big) \Big) = o_{\mathbb{P}}\Big(\| \boldsymbol{\beta}_n \|_2^{-2} \Big), \tag{1.7}$$

où $\mathcal{F}_M := \big\{ f\mathbb{1}_{\{F\leq M\}},\ f \in \mathcal{F} \big\}$ *et*

$$\overline{P}(\mathbf{p}, \mathbf{y}) := \sum_{i\geq 1} p_i \delta_{y_i},\ \textit{for } \mathbf{p} \in \mathbb{S},\ \mathbf{y} \in \mathfrak{X}^{\mathbb{N}}. \tag{1.8}$$

Alors

$$\mathbb{E}\big(\| G_n \|_{\mathcal{F}} \big) \to 0. \tag{1.9}$$

Notre second résultat est un théorème de Donsker. Comme, pour n fixé, les $X_{i,n}$ sont seulement conditionnellement i.i.d., l'approximation asymptotique de la loi des G_n ne sera pas faite pat des analogues gaussiens W_n des G_n, mais par des mélanges W_n des $\mathbf{P}_{n,\boldsymbol{\beta}_n}$-ponts browniens par les $\boldsymbol{\beta}_n$, autrement dit : pour tous $p \geq 1$, $\mathbf{f} := (f_1, \ldots, f_p) \in \mathcal{F}^p$ et $\mathbf{p} \in \mathbb{S}$, en écrivant $\mathbf{Q}^{\mathbf{f}}_{n,\mathbf{p}}$ pour la loi normale centré de matrice variance

$$\Sigma^{\mathbf{f}}_{n,\mathbf{p}} := \Big[\mathbf{P}_{n,\mathbf{p}}\Big((f_j - \mathbf{P}_{n,\mathbf{p}}(f_j))(f_{j'} - \mathbf{P}_{n,\mathbf{p}}(f_{j'})) \Big) \Big]_{(j,j')\in\{1,\ldots,p\}^2},$$

nous définissons, pour chaque borélien $A \subset \mathbb{R}^p$:

$$\mathbf{Q}^{\mathbf{f}}_n(A) := \mathbb{E}\Big(\mathbf{Q}^{\mathbf{f}}_{n,\boldsymbol{\beta}_n}(A) \Big).$$

Le théorème d'extension de Kolmogorov nous assure l'existence d'un loi de probabilité $\mathbb{P}'_n$ sur $\Omega' := \mathbb{R}^{\mathcal{F}}$, équippé de sa tribu Borélienne produit ($\mathbb{P}'_n$-complétée) $\mathcal{X}'$, la loi $\mathbb{P}'_n$ étant compatible avec le système projectif de lois $\big\{ \mathbf{Q}^{\mathbf{f}}_n,\ \mathbf{f} \in \mathcal{F}^p,\ p \geq 1 \big\}$. Nous définissons W_n comme l'application canonique sur $\big(\Omega', \mathcal{X}', \mathbb{P}'_n\big)$. Cette application n'est pas nécessairement mesurable par rapport aux boréliens de $\big(\ell^\infty(\mathcal{F}), \| \cdot \|_{\mathcal{F}} \big)$. Ce problème de mesurabilité est mis de côté en intriduisant des espérance extérieures (voir Van der Vaart and Wellner (1996, Chapter 1.2)). Pour simplifier les notations, nous adopterons la convention suivante : à chaque fois qu'une application $\mathfrak{h}$ est définie sur un espace probabilisé, le symbole $\mathbb{E}^*(\mathfrak{h})$ se rapportera à l'espérance éxtérieur relativement à cet espace probabilisé. Nous adopterons la même convention pour les probabilités extérieures $\mathbb{P}^*$.

Théorème 2 *Supposons que, pour tout* $n \geq 1$, $\| \boldsymbol{\beta}_n \|_2^2 = 1$ *presque sûrement, et que* $\| \boldsymbol{\beta}_n \|_4 \to 0$ *en probabilité. Supposons également que* $\mathbb{E}\Big(F^2(Y_{1,n})\Big) < \infty$ *pour tout* n*, et que :*

$$\lim_{M\to\infty} \overline{\lim_{n\to\infty}}\, \mathbb{E}\Big(F^2(Y_{1,n}) 1_{\{F(Y_{1,n})\geq M\}}\Big) = 0, \tag{1.10}$$

$$\int_0^\infty \sqrt{\log\Big(\sup_{Q\ probab.} N\big(\epsilon \| F \|_{Q,2}, \mathcal{F}, \| \cdot \|_{Q,2}\,\big)\Big)} d\epsilon < \infty. \tag{1.11}$$

Alors, pour tout $n \geq 1$, W_n *est presque sûrement borné, et* $\| W_n \|_{\mathcal{F}}$ *est Borel mesurable. Supposons de plus que, pour une semi métrique* ρ_0 *qui rend* $\mathcal{F}$ *précompact, on a*

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}}^{\mathbb{P}^*} \sup_{(f_1,f_2)\in\mathcal{F}^2, \rho_0(f_1,f_2)\leq\delta} \| f_1 - f_2 \|_{\mathbf{P}_{n,\boldsymbol{\beta}_n},2} = 0, \tag{1.12}$$

où $\mathbf{P}_n$ *désigne la loi de* $Y_{1,n}$ *et le symbole* $\overline{\lim}_{n\to\infty}^{\mathbb{P}^*}$ *désigne la limite sup en probabilité. Alors*

$$d_{BL}\Big(G_n, W_n\Big) := \sup_{B\in BL1} \Big|\mathbb{E}^*\Big(B\big(G_n\big)\Big) - \mathbb{E}^*\Big(B\big(W_n\big)\Big)\Big| \to 0, \tag{1.13}$$

où $BL1$ *est l'ensemble des fonctions réelles sur* $\big(\ell^\infty(\mathcal{F}), \| \cdot \|_{\mathcal{F}}\,\big)$*, 1-Lipschitziennes, et bornées par 1.*

Une conséquence directe du Théorème 2 est que, pour tout intervalle I, on a

$$\lim_{n\to\infty} \mathbb{P}\Big(\, \| G_n \|_{\mathcal{F}} \in I\Big) - \mathbb{P}'_n\Big(\, \| W_n \|_{\mathcal{F}} \in I\Big) = 0.$$

Nous conclurons cet exposé par une application, en statistique bayésienne non paramétrique, du Théorème 2 au comportement asymptotique de la loi à postériori d'une loi à priori de (processus de) Dirichlet pour la norme $\| \cdot \|_{\mathcal{F}}$. Étant donnée une loi de probabilité $\overline{\alpha}'$ sur $\big(\mathfrak{X}, \mathcal{X}\big)$, et $n \geq 1$, nous définissons l'application suivante sur $\mathfrak{X}^{\mathbb{N}}$: $\mathbb{T}_{\overline{\alpha}'}^{(n)}$: $(\mathbf{P}_i)_{i\geq 1} \to \Big(\int f d\mathbf{P}_n - \int f d\overline{\alpha}'\Big)_{f\in\mathcal{F}}$. Le symbole $\to_{\mathcal{L}}$ désigne ici la convergence en loi au sense de Hoffman-Jørgensen (voir (Van der Vaart and Wellner, 1996, Chapter 1.3)).

Corollaire 2.1 *Soit* $M > 0$ *et soit* $\overline{\alpha}$ *une mesure de probabilité sur* $\big(\mathfrak{X}, \mathcal{X}\big)$*. Soit* $(X_i)_{i\geq 1}$ *une suite i.i.d de variables aléatoires* $(\mathfrak{X}, \mathcal{X})$*, ayant pour loi* $\mathbf{P}_0$*. Supposons que* $\mathcal{F}$ *admet une enveloppe* $F \geq 1$ *qui soit de carré* $\overline{\alpha} + \mathbf{P}_0$ *integrable, et satisfaisant (1.11). Alors, pour presque toute réalisation* $(x_i)_{n\geq 1}$ *de* $(X_i)_{n\geq 1}$*, en écrivant* $\theta_n := M/(M+n)$*, on a , lorsque* $n \to \infty$ *:*

$$\sqrt{n}\mathbb{T}^{(n)}_{\theta_n\overline{\alpha}+(1-\theta_n)\overline{\alpha}_{(x_1,\ldots,x_n)}} \to_{\mathcal{L}} W_{\mathbf{P}_0},$$
$$\sqrt{n}\mathbb{T}^{(n)}_{\overline{\alpha}_{(x_1,\ldots,x_n)}} \to_{\mathcal{L}} W_{\mathbf{P}_0},$$

où ces deux suites sont considérées comme des éléments aléatoires sur $\Big(\mathfrak{X}^{\mathbb{N}}, \bigotimes_{n\geq 1} DP\big(M + n, \theta_n\overline{\alpha} + (1-\theta_n)\overline{\alpha}_{(x_1,\ldots,x_n)}\big)\Big)$.

Références

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230.

Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (2010). *Bayesian nonparametrics*, volume 28. Cambridge University Press.

Sethuraman, J. (1994). A constructive definition of dirichlet measures. *Statist. Sinica.*

Sheehy, A. and Wellner, J. A. (1992). Uniform donsker classes of functions. *Ann. Probab.*, pages 1983–2030.

Van der Vaart, A. and Wellner, J. (1996). *Weak convergence and empirical processes.* Springer.