

Estimation bayésienne non paramétrique de l'état et du bruit impulsif dans les systèmes dynamiques non linéaires

Nouha JAOUA^{1,2}, Emmanuel DUFLOS^{1,2}, Philippe VANHEEGHE^{1,2}, François SEPTIER^{1,3}

¹LAGIS UMR CNRS 8219, 59651 Villeneuve d'Ascq, France

²Ecole Centrale de Lille, 59651 Villeneuve d'Ascq, France

³Institut Mines-Télécom / Télécom Lille 1, 59658 Villeneuve d'Ascq, France

nouha.jaoua@ec-lille.fr, emmanuel.duflos@ec-lille.fr

philippe.vanheeghe@ec-lille.fr, francois.septier@telecom-lille1.eu

Résumé – Dans cet article, nous traitons le problème de l'estimation conjointe de l'état et de la densité de probabilité du bruit dans les systèmes dynamiques non linéaires en présence de bruit d'observation impulsif et multimodal. Un modèle non paramétrique bayésien basé sur les mélanges de processus de Dirichlet est introduit pour modéliser le bruit impulsif. Un algorithme de Monte Carlo séquentiel basé sur une loi d'importance efficace est ensuite proposé afin de réaliser l'estimation en ligne de l'état et de la densité du bruit. Les simulations effectuées montrent la validité et la robustesse de notre approche.

Abstract – In this paper, we address the problem of online state and measurement noise density estimation in nonlinear dynamic state-space models. We are especially interested in making inference in the presence of impulsive and multimodal noise. The proposed method relies on the introduction of a flexible Bayesian nonparametric noise model based on Dirichlet Process mixtures. A novel approach based on sequential Monte Carlo methods is proposed to perform the optimal online estimation. Simulation results demonstrate the efficiency and the robustness of this approach.

1 Introduction

Le domaine du traitement statistique du signal a été toujours dominé par l'hypothèse de gaussianité pour la modélisation du bruit. Cependant, dans un nombre croissant d'applications, les perturbations rencontrées s'éloignent fortement de ces modèles classiques. C'est en particulier le cas des bruits impulsifs que nous rencontrons dans de nombreux domaines, notamment celui des télécommunications. Un bruit de nature impulsive est caractérisé par une faible probabilité d'apparition mais de très fortes amplitudes. La modélisation de ce type de bruit par une loi gaussienne n'est pas pertinente à cause de sa grande variabilité. Il convient alors de modéliser ce type de bruit non pas par une gaussienne mais par une distribution qui n'admet pas forcément une variance finie. Dans le cadre des distributions non gaussiennes à variance infinie sont apparues les lois α -stables. Elles font partie d'une classe riche de distributions de probabilité qui englobent les lois de Gauss, de Cauchy et de Lévy et qui autorise l'asymétrie et les queues lourdes.

Les distributions α -stables n'ont pas de forme explicite générale pour leur densité de probabilité. Pour cette raison, elles sont généralement définies à partir de leur fonction caractéristique définie comme suit :

$$\varphi(t) = \begin{cases} \exp(i\mu t - \gamma |t|^\alpha [1 + i\beta \operatorname{sign}(t) \tan \frac{\alpha\pi}{2}]), & \alpha \neq 1 \\ \exp(i\mu t - \gamma |t|^\alpha [1 + i\beta \operatorname{sign}(t) \frac{2}{\pi} \log |t|]), & \alpha = 1 \end{cases}$$

Par conséquent, une distribution α -stable est totalement définie par quatre paramètres : α , β , γ et μ . Par convention, une distribution α -stable de paramètres α , μ , β , γ est notée $\mathcal{S}_\alpha(\mu, \beta, \gamma)$. Pour une description plus détaillée des distributions α -stable, on peut se référer au livre de Nikias et Shao [1].

Dans ce papier, nous nous intéressons au problème de l'estimation conjointe de l'état et de la densité du bruit dans les modèles dynamiques non linéaires. L'approche proposée repose sur l'introduction d'un modèle non paramétrique bayésien basé sur les mélanges de processus de Dirichlet (Dirichlet Process Mixtures, DPMs) pour modéliser le bruit α -stable comme un mélange infini de distributions de Cauchy. Un filtre particulaire basé sur des lois d'importances efficaces est ensuite implémenté pour réaliser l'estimation en ligne de l'état et de la densité du bruit d'observation.

2 Mélange de processus de Dirichlet pour l'estimation de la densité

On considère un ensemble d'observations z_1, \dots, z_n distribuées selon une densité de probabilité inconnue F . On désire estimer dans un cadre bayésien la densité de probabilité F à partir de la connaissance de l'ensemble d'observations. On s'intéresse tout particulièrement à la classe des densités de probabilité pou-

vant s'écrire sous le modèle de mélange suivant :

$$F(z) = \int_{\Theta} f(z|\theta) d\mathbb{G}(\theta) \quad (1)$$

avec $\theta \in \Theta$ une variable latente, $f(\cdot|\theta)$ la densité mélangée connue et \mathbb{G} la distribution de mélange. Cette dernière est supposée inconnue de distribution *a priori* $P(\mathbb{G})$. \mathbb{G} est alors appelée mesure de probabilité aléatoire (Random Probability Measure, RPM). Dans ce travail, nous supposons que \mathbb{G} est distribuée selon un processus de Dirichlet (Dirichlet Process, DP).

Ferguson [3] a introduit le DP comme une mesure de probabilité dans l'espace des mesures de probabilité. Il permet donc de définir, dans le cadre de l'estimation bayésienne, un *a priori* sur une distribution de probabilité inconnue. Un DP est défini par deux paramètres : une mesure de probabilité de base \mathbb{G}_0 dans un espace mesurable Θ et un paramètre d'échelle α . On le note par $DP(\mathbb{G}_0, \alpha)$. Les DPs admettent des propriétés très intéressantes qui les rendent particulièrement attractifs dans un cadre d'estimation car elles permettent la mise en oeuvre de mécanismes simplifiés des principes d'inférence bayésienne. Une propriété très importante est que les réalisations d'un DP sont discrètes avec une probabilité 1. Sethuraman a montré qu'une réalisation $\mathbb{G} \sim DP(\mathbb{G}_0, \alpha)$ d'un DP peut avoir la représentation stick-breaking suivante [2] :

$$\mathbb{G} = \sum_{k=1}^{\infty} \pi_k \delta_{U_k} \quad \text{avec} \quad U_k | \mathbb{G}_0 \sim \mathbb{G}_0, \quad k = 1, 2, \dots \quad (2)$$

avec $\theta_k \sim \mathbb{G}$, $\pi_k = \beta_k \prod_{j=1}^{k-1} (1 - \beta_j)$ et $\beta_k \sim \mathcal{B}(1, \alpha)$ où $\mathcal{B}(\cdot, \cdot)$ est la loi bêta. En utilisant l'équation (1), la distribution inconnue F peut s'écrire comme suit :

$$F(z) = \sum_{k=1}^{\infty} \pi_k f(z|\theta_k) \quad (3)$$

Le caractère discret de la distribution G permet l'estimation directe du nombre de composantes sans avoir recours à d'autres techniques telles que les sauts réversibles.

Une autre motivation fondamentale pour l'utilisation du DP comme RPM est la représentation en urne de Polya de la distribution prédictive qui peut être obtenue en marginalisant selon la RPM \mathbb{G} [4] :

$$\theta_{n+1} | \theta_{1:n} \sim \frac{\alpha}{\alpha + n} \mathbb{G}_0 + \frac{1}{\alpha + n} \sum_{i=1}^{n-1} \delta_{\theta_i} \quad (4)$$

Cette représentation est très utile en pratique car elle permet d'échantillonner selon un DP sans la construction explicite de la RPM.

En choisissant comme distribution *a priori* pour la RPM \mathbb{G} un DP, on peut reformuler le problème d'estimation selon le modèle hiérarchique suivant connu sous le nom de DPM :

$$\begin{aligned} \mathbb{G} &\sim DP(\mathbb{G}_0, \alpha), \\ \theta_i | \mathbb{G} &\sim \mathbb{G}, \\ z_i | \theta_i &\sim f(\cdot | \theta_i) \end{aligned}$$

3 Modélisation du bruit avec un DPM de Cauchy

Nous considérons le modèle dynamique non linéaire suivant :

$$\begin{cases} x_{t+1} = g_t(x_t, w_t) \\ y_t = h_t(x_t, v_t) \end{cases} \quad (5)$$

où x_t est le vecteur d'état caché, y_t est l'observation, g_t et h_t sont les fonctions connues d'évolution de l'état et d'observation, w_t est le bruit d'état et v_t est le bruit d'observation. Nous supposons que la densité de probabilité du bruit w_t est connue. Le bruit v_t est supposé être α -stable avec une densité de probabilité inconnue pouvant être multimodale et asymétrique.

Dans ce papier, nous supposons que la densité de probabilité du bruit d'observation est un DPM de distributions de Cauchy. Dans ce cas, la densité mélangée $f(\cdot|\theta_t)$ est supposée être une distribution de Cauchy, notée $\mathcal{C}(a_t, m_t)$. Nous choisissons une loi inverse-gamma $\mathcal{IG}(a^a, b^a)$ comme loi *a priori* pour a_t et une loi normale $\mathcal{N}(m^m, \Sigma^m)$ comme loi *a priori* pour m_t . La distribution de base \mathbb{G}_0 est définie comme étant le produit des lois *a priori* des paramètres de la densité mélangée :

$$\mathbb{G}_0 = \mathcal{IG}(a^a, b^a) \times \mathcal{N}(m^m, \Sigma^m)$$

En résumé, nous obtenons le modèle hiérarchique suivant :

$$\begin{aligned} \mathbb{G} | \Phi &\sim DP(\mathbb{G}_0, \alpha_{DPM}) \\ \theta_t | \mathbb{G} &\sim \mathbb{G} \\ v_t | \theta_t &\sim \mathcal{C}(a_t, m_t) \end{aligned} \quad (6)$$

où $\Phi = \{\alpha_{DPM}, a^a, b^a, m^m, \Sigma^m\}$ et $\theta_t = \{a_t, m_t\}$ sont respectivement l'ensemble des hyperparamètres et la variable latente donnant à chaque instant t les paramètres de la distribution de Cauchy. Dans ce travail, les hyperparamètres sont supposés fixes et connus. Ce modèle peut s'écrire de façon équivalente $v_t \sim F(v_t)$ où $F(v_t)$ est la densité de probabilité du bruit d'observation définie comme suit :

$$F(v_t) = \int f(\cdot | \theta_t) d\mathbb{G}(\theta_t)$$

4 Filtre particulaire pour l'estimation de l'état et de la densité du bruit

Etant donné le modèle statistique défini par les équations (5)-(6), notre objectif est l'estimation conjointe de l'état x_t et de la variable latente θ_t conditionnellement aux observations $y_{1:t}$. Dans un cadre bayésien, ceci revient à calculer la densité *a posteriori* $p(x_t, \theta_t | y_{1:t}, \Phi)$. Comme son calcul analytique n'est pas possible, nous proposons de l'approcher par filtrage particulaire. En effet, cette densité est approchée à l'aide d'un système de N particules pondérées $\left\{ (x_t^{(i)}, \theta_t^{(i)}, \omega_t^{(i)}) \right\}_{i=1}^N$:

$$p(x_t, \theta_t | y_{1:t}, \Phi) \simeq \sum_{i=1}^N \omega_t^{(i)} \delta_{x_t^{(i)}, \theta_t^{(i)}}(x_t, \theta_t) \quad (7)$$

où δ est le fonction de Dirac, $x_t^{(i)}$ et $\theta_t^{(i)}$ sont échantillonnées selon la densité d'importance $q(x_t, \theta_t | x_{0:t-1}, \theta_{1:t-1}, y_{1:t}, \Phi)$ et $\omega_t^{(i)}$ est le poids normalisé associé à la $i^{\text{ème}}$ particule.

Afin de garantir une efficacité maximale de notre filtre particulaire, nous proposons d'utiliser la densité d'importance optimale qui minimise la variance des poids

$$q(x_t, \theta_t | x_{0:t-1}, \theta_{1:t-1}, y_{1:t}, \Phi) = p(x_t, \theta_t | x_{0:t-1}, \theta_{1:t-1}, y_t, \Phi) \quad (8)$$

L'échantillonnage des particules $x_t^{(i)}$ et $\theta_t^{(i)}$ à partir de l'équation (8) nécessite une expression analytique de la densité d'importance optimale. Cependant, le calcul analytique de cette densité n'est pas possible. En utilisant la règle de Bayes, cette densité peut s'écrire :

$$q(x_t, \theta_t | x_{0:t-1}, \theta_{1:t-1}, y_{1:t}, \Phi) \propto p(y_t | x_{0:t}, \theta_{1:t}, \Phi) p(x_t | \theta_{1:t}, x_{0:t-1}) p(\theta_t | \theta_{1:t-1}, \Phi) \quad (9)$$

Par conséquent, une approximation de cette densité peut être obtenue en utilisant l'intégration de Monte Carlo et l'échantillonnage d'importance. Pour ce faire, nous considérons, pour chaque couple de particule $(x_t^{(i)}, \theta_t^{(i)})$, un système de N_{IS} particules auxiliaires $\{(\check{x}_{t,i}^{(j)}, \check{\theta}_{t,i}^{(j)})\}_{j=1}^{N_{IS}}$ où les particules du vecteur d'état $\check{x}_{t,i}^{(j)}$ sont échantillonnées selon la loi d'évolution de l'état :

$$\check{x}_{t,i}^{(j)} \sim p(x_t | x_{t-1}^{(i)}) \quad (10)$$

et les particules de la variable latente $\check{\theta}_{t,i}^{(j)}$ sont échantillonnées selon la loi de prédiction admettant la représentation en urne de Polya (cf. section 2) :

$$\check{\theta}_{t,i}^{(j)} | \theta_{1:t-1}^{(i)} \sim \frac{\alpha}{\alpha + t} \mathbb{G}_0 + \frac{1}{\alpha + t} \sum_{k=1}^{t-1} \delta_{\theta_k^{(i)}}(\theta_k) \quad (11)$$

La densité d'importance optimale est alors approchée à l'aide de ce système de particules pondérées par la densité empirique

$$p(x_t, \theta_t | x_{0:t-1}, \theta_{1:t-1}, y_{1:t}, \Phi) \simeq \sum_{j=1}^{N_{IS}} \frac{\check{\omega}_{t,i}^{(j)}}{S_{\check{\omega}}} \delta_{\check{x}_{t,i}^{(j)}, \check{\theta}_{t,i}^{(j)}}(x_t, \theta_t) \quad (12)$$

où $\check{\omega}_{t,i}^{(j)}$ est le poids non normalisé associé à la $j^{\text{ème}}$ particule auxiliaire donné par

$$\check{\omega}_{t,i}^{(j)} = p(y_t | \check{x}_{t,i}^{(j)}, \check{\theta}_{t,i}^{(j)}, \Phi) \quad (13)$$

et $S_{\check{\omega}}$ est la somme des poids non normalisés $S_{\check{\omega}} = \sum_{j=1}^{N_{IS}} \check{\omega}_{t,i}^{(j)}$.

Pour générer le $i^{\text{ème}}$ couple de particules $(x_t^{(i)}, \theta_t^{(i)})$ selon cette approximation de la densité d'importance optimale conjointe, il suffit de tirer une particule à partir du système $\{(\check{x}_{t,i}^{(j)}, \check{\theta}_{t,i}^{(j)})\}_{j=1}^{N_{IS}}$

en utilisant les poids $\check{\omega}_{t,i}^{(1)}, \check{\omega}_{t,i}^{(2)}, \dots, \check{\omega}_{t,i}^{(N_{IS})}$ comme probabilités de sélection. Une façon de faire est de générer une variable aléatoire J à valeurs dans $\{1, 2, \dots, N_{IS}\}$ en utilisant la loi multinomiale :

$$J \sim \text{Multinomial} \left(\frac{\check{\omega}_{t,i}^{(1)}}{S_{\check{\omega}}}, \frac{\check{\omega}_{t,i}^{(2)}}{S_{\check{\omega}}}, \dots, \frac{\check{\omega}_{t,i}^{(N_{IS})}}{S_{\check{\omega}}} \right) \quad (14)$$

Le $i^{\text{ème}}$ couple de particules est alors donné par :

$$(x_t^{(i)}, \theta_t^{(i)}) = (\check{x}_{t,i}^{(j=J)}, \check{\theta}_{t,i}^{(j=J)}) \quad (15)$$

Avec une telle densité d'importance, nous obtenons l'expression suivante pour la mise à jour des poids :

$$\omega_t^{(i)} \propto \omega_{t-1}^{(i)} p(y_t | x_{0:t-1}^{(i)}, \theta_{1:t-1}^{(i)}, \Phi) \quad (16)$$

En marginalisant selon x_t et θ_t , la densité $p(y_t | x_{0:t-1}^{(i)}, \theta_{1:t-1}^{(i)}, \Phi)$ peut s'écrire :

$$p(y_t | x_{0:t-1}, \theta_{1:t-1}, \Phi) = \int p(y_t | x_t, \theta_t) p(x_t | \theta_{1:t}, x_{0:t-1}) p(\theta_t | \theta_{1:t-1}, \Phi) dx_t d\theta_t \quad (17)$$

Cette densité peut être approchée par intégration de Monte Carlo en utilisant le système de particules auxiliaires $\{(\check{x}_{t,i}^{(j)}, \check{\theta}_{t,i}^{(j)})\}_{j=1}^{N_{IS}}$. En procédant ainsi, cette densité est donnée par la somme des poids non normalisés $S_{\check{\omega}}$.

Le filtre particulaire proposé pour l'estimation conjointe de l'état et de la densité du bruit d'observation est résumé dans l'algorithme 1.

Initialisation
for $i = 1$ **a** N **do**
 Échantillonner $x_0^{(i)} \sim p_0(x_0)$;
 Échantillonner $\theta_0^{(i)} \sim \mathbb{G}_0$;
 Initialiser $\omega_0^{(i)} = 1/N$;
end
pour $t = 1$ **a** T **faire**
 pour $i = 1$ **a** N **faire**
 pour $j = 1$ **a** N_{IS} **faire**
 Échantillonner $\check{x}_{t,i}^{(j)} \sim p(x_t | x_{0:t-1}^{(i)})$ avec (10) ;
 Échantillonner $\check{\theta}_{t,i}^{(j)} \sim p(\theta_t | \theta_{0:t-1}^{(i)}, \Phi)$ avec (11) ;
 Calculer les poids : $\check{\omega}_{t,i}^{(j)}$ avec (13) ;
 fin
 Calculer : $S_{\check{\omega}} = \sum_{j=1}^{N_{IS}} \check{\omega}_{t,i}^{(j)}$;
 Normaliser les poids : $\check{\omega}_{t,i}^{(j)} = \check{\omega}_{t,i}^{(j)} / S_{\check{\omega}}$;
 Échantillonner
 $J \sim \text{Multinomial}(\check{\omega}_{t,i}^{(1)}, \check{\omega}_{t,i}^{(2)}, \dots, \check{\omega}_{t,i}^{(N_{IS})})$;
 Faire $x_t^{(i)} = \check{x}_{t,i}^{(J)}$ et $\theta_t^{(i)} = \check{\theta}_{t,i}^{(J)}$;
 Calculer les poids : $\omega_t^{(i)} \propto \omega_{t-1}^{(i)} S_{\check{\omega}}$;
 fin
 Normaliser les poids
 $\omega_t^{(i)} = \omega_t^{(i)} / \sum_{j=1}^N \omega_t^{(j)}$, $i = 1, \dots, N$;
 si $N_{eff} < \eta$ **alors** Ré-échantillonnage ;
fin

Algorithme 1: Filtre particulaire pour l'estimation en ligne de l'état et de la densité du bruit d'observation

5 Exemple illustratif

La performance du filtre particulaire proposé est évaluée en utilisant le modèle dynamique fortement non linéaire suivant :

$$\begin{cases} x_{t+1} = 0.5x_t + 25 \frac{x_t}{1+x_t^2} + 8 \cos(1.2(t+1)) + w_t \\ y_t = \frac{x_t^2}{20} + v_t \end{cases}$$

Ce modèle a été simulé avec les paramètres suivant : $T = 300$, $w_t \sim \mathcal{N}(0, 10)$ et $v_t \sim 0.6\mathcal{S}_{1.2}(0, 2, -5) + 0.4\mathcal{S}_{1.5}(0.5, 1.5, 5)$. Les hyperparamètres de la distribution de base sont fixés à $a^a = 5$, $b^a = 4$, $m^m = 0$ et $\Sigma^m = 50$. Le filtre particulaire proposé est utilisé avec $N = 200$ particules et $N_{IS} = 100$ particules auxiliaires.

Les résultats obtenus sont illustrés dans les figures 1, 2 et 3. Les états réel et estimé ainsi que les observations sont tracés dans la figure 1 (haut). L'évolution en fonction du temps de la différence entre l'état réel et l'état estimé est aussi illustrée sur la figure 1 (bas). La figure 2 montre la densité réelle ainsi que la densité estimée du bruit d'observation. D'après ces figures, nous pouvons remarquer la capacité de filtrage des pics liés à la présence du bruit impulsif ainsi que la bonne approximation de la densité du bruit d'observation. La courbes d'évolution en fonction du temps de la distance de Kullback-Leibler entre la densité réelle et la densité estimée est tracée sur la figure 3. À partir de cette figure, nous pouvons constater la convergence progressive de la densité estimée vers la densité réelle.

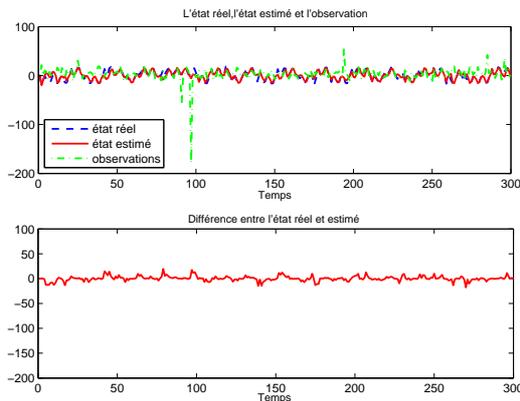


FIGURE 1 – Estimation de l'état. (Haut) Etat réel, état estimé et l'observation. (Bas) Évolution temporelle de l'erreur entre l'état réel et l'état estimé.

6 Conclusion

Dans cet article, nous avons présenté une approche non paramétrique bayésienne pour l'estimation conjointe de l'état et de la densité du bruit dans les modèles dynamiques non linéaires en présence d'un bruit d'observation α -stable asymétrique et multimodal. L'originalité de ce travail consiste à modéliser un processus α -stable par un mélange de processus de Dirichlet de distributions de Cauchy. Les simulations effectuées ont montré que les mélanges de processus de Dirichlet de distributions

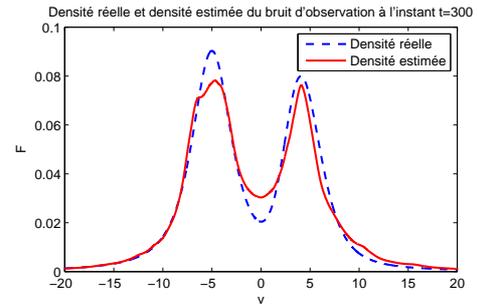


FIGURE 2 – Estimation de la densité du bruit à l'instant t=300.

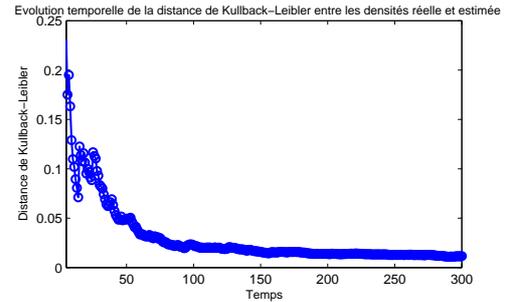


FIGURE 3 – Évolution temporelle de la distance de Kullback-Leibler entre la densité réelle et la densité estimée.

de Cauchy sont très appropriés pour la modélisation des processus α -stables. Ce résultat est très intéressant vu qu'il permet de surmonter les difficultés souvent rencontrées lors de l'inférence en l'absence d'une expression analytique de la densité de probabilité des distributions α -stables. De plus, dans le cadre des modèles dynamiques non linéaires, nous avons montré que l'estimation de l'état est tout à fait possible même en présence d'un bruit extrêmement impulsif ayant des statistiques inconnues. Finalement, il faut noter que l'approche proposée n'est pas limitée au cas du bruit α -stable et peut être appliquée avec d'autres types de bruit. Dans des travaux futurs, nous envisageons de supposer que les hyperparamètres de la distribution de base sont inconnus et de les inclure dans l'inférence.

Références

- [1] C.L. Nikias et M. Shao. *Signal Processing with α -stable Distributions and Applications*. John Wiley and Sons, 1995.
- [2] J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4 :639-650, 1994
- [3] T.S. Ferguson. A Bayesian Analysis of some Nonparametric Problems. *The Annals of Statistics*, 1(2) :209-230, 1973.
- [4] D. Blackwell et J.B. Macqueen. Ferguson Distributions via Polya Urn Schemes. *The Annals of Statistics*, 1 :353-355, 1973