Démonstrations - SMC

Demonet Guillaume

May 2016

RETOUR AU NOTEBOOK

Application de la méthode de l'entropie croisée

On cherche à estimer la quantité :

$$\mathbb{E}_\theta[H(\mathbf{X})] = \int H(\mathbf{x})f(\mathbf{x}|\theta)\mathrm{d}\mathbf{x} \tag{1}$$

avec H une fonction objectif et $f(\cdot|\theta)$ une densité de probabilité définie par ses paramètres θ.

Dans notre cas d'application, la quantité à estimer sera nommée $p(n,t)$.
Le processus $\mathbf{X}$ considéré est le processus des temps d'inter-arrivées et des temps de service. Un tel processus définit toujours la trajectoire d'une file d'attente à simple guichet, que l'on peut noter $\mathbf{Y}$ comme le processus des états (ou longueurs) successifs de la file. On a donc ϕ telle que $\mathbf{Y} = \phi(\mathbf{X})$ est surjective. L'évènement où la longueur de la file a dépassé un seuil $\mathbf{n}$ au cours de la période $[0,t]$, se traduisant par l'ensemble

$$B = \{(Y_k)_{[0,t]} \quad | \quad \exists k, Y_k \geq n\}$$

est "accessible" avec le processus $\mathbf{X}$:

$$\mathbf{X} \in A := \phi^{-1}(B) \Leftrightarrow \mathbf{Y} \in B$$

(attention l'équivalence est seulement vraie parce que $\mathbf{Y} = \phi(\mathbf{X})$).

On a ainsi, pour revenir à (1) :

$$\begin{aligned} H(\mathbf{X}) &= \mathbb{1}_A(\mathbf{X}), \\ \theta &= (\lambda_1, \lambda_2) \end{aligned}$$

La densité jointe du processus $\mathbf{X}$ est alors telle que, en notant m_a le nombre d'arrivées ayant eu lieu, m_d le nombre de départs, et $\mathbf{X_a}$, $\mathbf{X_d}$ les sous-processus des temps de service et d'inter-arrivées respectivement :

$$f(\mathbf{X}|\theta) = \prod_{i=1}^{m_a} \lambda_1 \exp(-\lambda_1 \mathbf{X}_{\mathbf{a},i}) \times \prod_{j=1}^{m_d} \lambda_2 \exp(-\lambda_2 \mathbf{X}_{\mathbf{d},j}) \tag{2}$$

Le problème d'estimation que nous venons de décrire peut se résoudre en utilisant un estimateur convergent naïf CMC (pour *Crude Monte Carlo*) :

$$p_{CMC}(n,t) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_A(\mathbf{X}^i)$$

ce que nous avons implémenté en première partie (les $\mathbf{X}^i$ sont tirés selon $f(\cdot|\theta)$).

Cependant, lorsque A est rare, une estimation correcte de $p(n,t)$ grâce à l'estimateur CMC nécessite de grandes valeurs de N. La méthode de l'*Importance Sampling* permet de reformuler le problème à notre avantage.
On considère g un changement de mesure tel que

$$g(\mathbf{X}) = 0 \Rightarrow \mathbb{1}_A(\mathbf{X}) \cdot f(\mathbf{X}|\theta) = 0$$

Alors on a :

$$\begin{aligned} p(n,t) &= \mathbb{E}_\theta[\mathbb{1}_A(\mathbf{X})] \\ &= \int \mathbb{1}_A(\mathbf{x}) \frac{f(\mathbf{x}|\theta)}{g(\mathbf{x})} g(\mathbf{x}) \mathrm{d}\mathbf{x} \\ &= \mathbb{E}_g[\mathbb{1}_A(\mathbf{X}) W_{f,g}(\mathbf{X})] \end{aligned} \tag{3}$$

Ceci permet de construire un estimateur empirique de l'espérance en tirant les $\mathbf{X}^i$ selon g, en notant $W_{f,g}(\mathbf{X}) = \frac{f(\mathbf{X}|\theta)}{g(\mathbf{X})}$ le rapport de vraisemblance entre $f(\cdot|\theta)$ et g en $\mathbf{X}$. En pratique, on cherchera g dans la même famille que f, et l'on notera

$$W(\mathbf{X}|\theta_1, \theta_2) = \frac{f(\mathbf{X}|\theta_1)}{f(\mathbf{X}|\theta_2)}$$

La densité g^* optimale pour estimer p grâce à l'*Importance Sampling* est naturellement :

$$g^*(x) = \frac{\mathbb{1}_A(\mathbf{X}) \cdot f(\mathbf{X}|\theta)}{p} \tag{4}$$

Cependant, on ne peut pas la calculer puisque l'on cherche à **estimer** p. La méthode de l'entropie croisée consiste alors à chercher ϑ tel que l'entropie croisée (ou distance de Kullback-Leibler) entre g^* et $f(\cdot|\vartheta)$ soit minimale. Elle est définie ainsi :

$$\begin{aligned} \mathcal{D}(g^*, f_\vartheta) &= \mathbb{E}_{g^*} \ln \frac{g^*(\mathbf{X})}{f_\vartheta(\mathbf{X})} \\ &= \int g^*(\mathbf{x}) \ln g^*(\mathbf{x}) \mathrm{d}\mathbf{x} - \int g^*(\mathbf{x}) \ln f_\vartheta(\mathbf{x}) \mathrm{d}\mathbf{x} \end{aligned}$$

Minimiser cette distance revient à maximiser $\int g^*(\mathbf{x})\ln f_\vartheta(\mathbf{x})\mathrm{d}\mathbf{x}$. En remplaçant g^* par sa définition (voir (4)), on obtient :

$$\begin{aligned}\int g^*(\mathbf{x})\ln f_\vartheta(\mathbf{x})\mathrm{d}\mathbf{x} &= \int \frac{\mathbb{1}_A(\mathbf{X})\cdot f(\mathbf{X}|\theta)}{p}\ln f_\vartheta(\mathbf{x})\mathrm{d}\mathbf{x} \\ &= \frac{1}{p}\mathbb{E}_\theta[\mathbb{1}_A(\mathbf{X})\ln f_\vartheta(\mathbf{X})] \\ &=: \frac{1}{p}D(\vartheta)\end{aligned}$$

Il nous suffit alors de maximiser $D(\vartheta)$ pour trouver le ϑ optimal de *IS*. Cependant, utiliser directement un gradient en ϑ sur l'estimateur naïf de cette espérance ne permet pas de contourner le problème de la rareté de A. Il s'agit d'appliquer à nouveau la méthode *IS* à cette quantité, et d'utiliser le changement de paramètres pour obtenir une relation de récurrence entre θ_j et θ_{j+1} (comme expliqué dans A Tutorial on the Cross-Entropy Method). On dérive ainsi cette relation :

$$\theta_{j+1} = \underset{\vartheta}{\operatorname{argmax}} \sum_{i=1}^{N} \mathbb{1}_A(\mathbf{X}^i)\cdot W(\mathbf{X}^i|\theta,\theta_j)\cdot \ln f_\vartheta(\mathbf{X}^i) \tag{5}$$

où les $\mathbf{X}^i$ sont tirés selon $f(\cdot|\theta_j)$. On peut appliquer un gradient en ϑ sur la somme à maximiser (qui est bien convexe, grâce aux densités exponentielles), et l'annuler pour obtenir une formule explicite de θ_{j+1}. En remplaçant dans notre situation θ par (λ_1, λ_2), on obtient les relations :

$$\begin{aligned}\lambda_{1,j+1} &= \frac{\sum_{i=1}^{N}\mathbb{1}_A(\mathbf{X}^i)\cdot W(\mathbf{X}^i|\theta,\theta_j)\cdot m_a^i}{\sum_{i=1}^{N}[\mathbb{1}_A(\mathbf{X}^i)\cdot W(\mathbf{X}^i|\theta,\theta_j)\sum_{k=1}^{m_a^i}\mathbf{X_a}_{,k}^{\,i}]} \\ \lambda_{2,j+1} &= \frac{\sum_{i=1}^{N}\mathbb{1}_A(\mathbf{X}^i)\cdot W(\mathbf{X}^i|\theta,\theta_j)\cdot m_d^i}{\sum_{i=1}^{N}[\mathbb{1}_A(\mathbf{X}^i)\cdot W(\mathbf{X}^i|\theta,\theta_j)\sum_{k=1}^{m_d^i}\mathbf{X_d}_{,k}^{\,i}]}\end{aligned} \tag{6}$$

Le processus itératif défini ainsi est supposé convergent vers le ϑ optimal qui minimise $\mathcal{D}(g^*, f_\vartheta)$.

RETOUR AU NOTEBOOK