fuzzy_sequ_0504
Sur l'estimation de répartitions discrètes

J.-P. Laedermann *

050425, corr 050503

Table des matières

1 Introduction **1**

2 Ingrédients **2**
2.1 Eléments de la théorie bayésienne de la décision statistique 2
2.2 Hypersphères 2

3 Estimateurs des probabilités pour des situations connues **3**

4 Séquences aléatoires auto-consistantes **4**

5 Généralisation aux situations floues **5**

6 Conclusion **7**

7 Résumé **7**
7.1 Vu par $\psi \in \mathbb{S}^{n-1}$ 7
7.2 Vu par $i \in \{0 \dots n-1\}^{\mathbb{N}}$ 8

1 Introduction

Soit un ensemble fini de n situations pouvant apparaître selon une répartition $(p_k)_{k=0}^{n-1}$. La détermination de la probabilité de chaque situation se fait généralement en répétant l'expérience de manière indépendante et en utilisant la définition opérationnelle de Laplace pour obtenir les estimateurs

$$\widehat{p}_k = \frac{\text{Nombre d'apparitions de la situation k}}{\text{Nombre de répétitions de l'expérience}}$$

De plus, le théorème central limite permet d'obtenir une estimation de l'écart séparant la vraie valeur de l'estimation.

Cette méthode empirique fonctionne généralement assez bien, mais pose quelques problèmes. D'une part, cet estimateur et son incertitude sont tributaires du nombre de répétitions, donnant de bons résultats

*jean-pascal.laedermann@chuv.ch

seulement lorsque que ce nombre est *assez grand*, notion assez imprécise. D'autre part, il est difficile de généraliser cette définition lorsque l'observation de la situation est floue. Or la détermination d'une situation se fait souvent par l'intermédiaire d'un instrument de mesure donnant un résultat bruité.

Le but de ce travail est de baser l'estimation des probabilités associées à des évènements sur le théorème de Bayes, en partant des a priori non-informatifs de Jeffreys.

2 Ingrédients

2.1 Eléments de la théorie bayésienne de la décision statistique

Cette théorie distingue tout d'abord deux espaces, celui des états θ et celui des observations x. Chaque observation est supposée faite en présence d'un état inconnu. Le *modèle de mesure* est la relation stochastique liant l'état au résultat.de la mesure. Pour un état donné, on donne une probabilité de transition[1] $\mathrm{p}(x \mid \theta)$, qui donne la répartition des observations x étant donné θ. Moyennant une répartition *a priori* $\mathrm{p}(\theta)$ sur les états et une observation x, le théorème de Bayes donne la répartition *a posteriori* sur les états

$$\mathrm{p}(\theta \mid x) = \frac{\mathrm{p}(x \mid \theta)\mathrm{p}(\theta)}{\int d\theta \mathrm{p}(x \mid \theta)\mathrm{p}(\theta)}$$

Lorsque l'espace des états est une variété différentiable et que le modèle de mesure est deux fois différentiable par rapport à θ, on définit un a priori dit *de Jeffreys* de la manière suivante. On forme tout d'abord la *matrice d'information de Fisher*

$$I_{ij}(\theta) = -\int dx \mathrm{p}(x \mid \theta) \frac{\partial^2 \ln \mathrm{p}(x \mid \theta)}{\partial\theta_i \partial\theta_j}$$

L'a priori de Jeffreys est défini ensuite comme

$$\mathrm{p}_J(\theta) d\theta = \sqrt{\det I(\theta)} d\theta$$

Une propriété fondamentale de cette répartition est son invariance. En effet, on peut montrer que $\mathrm{p}_J(\theta)d\theta$ ne dépend pas du choix du système de coordonnées.

2.2 Hypersphères

Les estimateurs construits dans ce travail font usage des hypersphères unité. Rappelons quelques unes des leurs propriétés.

Dans l'espace euclidien $\mathbb{R}^n$, l'hypersphère unité de dimension $n-1$ est la variété différentiable

$$\mathbb{S}^{n-1} = \{\psi \in \mathbb{R}^n \mid \sum_k \psi_k^2 = 1\}$$

L'usage de la lettre ψ sera justifié plus bas. Les coordonnées sphériques usuelles sont donnés par

$$\begin{aligned}
\psi_0 &= \sin\theta_{n-2}\sin\theta_{n-3}\ldots\sin\theta_1\sin\theta_0 \\
\psi_1 &= \sin\theta_{n-2}\sin\theta_{n-3}\ldots\sin\theta_1\cos\theta_0 \\
\psi_2 &= \sin\theta_{n-2}\sin\theta_{n-3}\ldots\cos\theta_1 \\
\ldots & \\
\psi_{n-1} &= \cos\theta_{n-2}
\end{aligned} \tag{1}$$

[1] L'utilisation de la fonte roman p designera une probabilité générique, les espaces sous-jacents étant identifiables par les variables utilisées.

et la mesure invariante sous $SO(n)$ normalisée est donnée par

$$d\mu(\psi) = \frac{1}{A_{n-1}} \sin^{n-2}\theta_{n-2} \sin^{n-3}\theta_{n-3} \ldots \sin\theta_1 d\theta_0 d\theta_1 \ldots d\theta_{n-2} \tag{2}$$

$$A_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)} \tag{3}$$

$$\Gamma(1/2) = \sqrt{\pi} \tag{4}$$

Soit la suite d'entiers positifs $m = (m_0 \ldots m_{n-1})$. Les moments pairs relatifs à cette mesure sont données par

$$H(m) = \int \psi_0^{2m_0} \ldots \psi_{n-1}^{2m_{n-1}} d\mu(\psi) \tag{5}$$

On peut monter que la fonction H ci-dessus est une généralisation de la fonction bêta

$$H(m) = \frac{\Gamma(n/2)}{\pi^{n/2}} \frac{\prod_k \Gamma(m_k + 1/2)}{\Gamma(\sum_k (m_k + 1/2))} \tag{6}$$

Le facteur de normalisation est choisi de sorte que $H(0) = 1$

3 Estimateurs des probabilités pour des situations connues

Le problème est l'estimation d'une répartition p_k. Cette répartition peut être vue comme un point du simplexe de dimension $n-1$

$$\Delta^{n-1} = \{p \mid 0 \leqslant p_k \leqslant 1, \sum_k p_k = 1\}$$

Ce simplexe sera l'espace des états pour un modèle de mesure dont les observations sont les situations. D'une façon canonique, la probabilité d'obtenir la situation k étant donné que la répartition est p est évidemment

$$\mathrm{p}(k \mid p) = p_k$$

Ce modèle de mesure canonique est deux fois différentiable par rapport p. On peut donc calculer l'a priori de Jeffreys associé $\mathrm{p}_J(p)$. Il est dit plus haut que cette répartition est indépendante de la paramétrisation du simplexe. Choisissons les nouvelles coordonnées

$$\psi_k = \sqrt{p_k} \tag{7}$$

Comme $\sum_k \psi_k^2 = \sum_k p_k = 1$, on obtient un modèle de mesure sur un espace des états qui est la partie positive de l'hypersphère $\mathbb{S}^{n-1}$

$$\mathrm{p}(k \mid \psi) = \psi_k^2$$

Or il se trouve que **l'a priori de Jeffreys pour ce modèle est la mesure invariante** μ citée sopra :

$$d\mathrm{p}_J(\psi) = d\mu(\psi) \tag{8}$$

Ce résultat permet d'appliquer le théorème de Bayes à une suite d'observations indépendantes. Soit

$$i = (i_0 \ldots i_{t-1})$$

une telle suite de longueur t, on obtient l'a posteriori sur la sphère $\mathbb{S}^{n-1}$ étant donné i

$$d\mathrm{p}(\psi \mid i) = \frac{\psi_{i_0}^2 \dots \psi_{i_{t-1}}^2 d\mu(\psi)}{\int \psi_{i_0}^2 \dots \psi_{i_{t-1}}^2 d\mu(\psi)} \tag{9}$$

Introduisant la fonction m donnant la multiplicité d'un indice k dans la suite i :

$$m(i)_k = \mathrm{Card}\{s \mid i_s = k\}$$

on obtient (sous-entendant la suite i pour simplifier la notation)

$$d\mathrm{p}(\psi \mid i) = \frac{\prod_k \psi_k^{2m_k} d\mu(\psi)}{H(m(i))} \tag{10}$$

La probabilité d'apparition d'une situation k peut être estimée par la moyenne de l'a posteriori,

$$\widehat{p}_k = \mathbb{E}(p_k) = \frac{\int \psi_k^2 \psi_{i_0}^2 \dots \psi_{i_{t-1}}^2 d\mu(\psi)}{H(m(i))} \tag{11}$$

Les incertitudes associées seront données par le biais de la matrice de covariance

$$\widehat{p_k p_l} = \mathbb{E}(p_k p_l) = \frac{\int \psi_k^2 \psi_l^2 \psi_{i_0}^2 \dots \psi_{i_{t-1}}^2 d\mu(\psi)}{H(m(i))} \tag{12}$$

Definisons par e_k une suite nulle d'entiers sauf le k^e dont la valeur est 1. On voit aisément que

$$\widehat{p}_k = \frac{H(m(i)+e_k)}{H(m(i))} \tag{13}$$

$$\widehat{p_k p_k} = \frac{H(m(i)+e_k+e_l)}{H(m(i))} \tag{14}$$

Le fait que $\sum_h m_h = t$ et Les propriétés de la fonction Gamma donnent immédiatement

$$\widehat{p}_k = \frac{m_k + 1/2}{t + n/2} \tag{15}$$

$$\widehat{p_k^2} = \frac{(m_k+1/2)(m_k+3/2)}{(t+n/2)(t+n/2+1)} \tag{16}$$

$$\widehat{p_k p_l} = \frac{(m_k+1/2)(m_l+1/2)}{(t+n/2)(t+n/2+1)} \tag{17}$$

4 Séquences aléatoires auto-consistantes

Soit une suite i de longueur $t-1$, la relation (15) donne donc les estimations bayesiennes pour les p_k si l'on a observé la suite i, que l'on notera aussi

$$\widehat{p}_k = \mathrm{p}_J(k \mid i_{t-2}, i_{t-3} \dots i_0)$$

A t, tirons au hasard selon ces p_k une nouvelle situation i_{t-1}. Formons une nouvelle suite par concaténation : (i, i_{t-1}). Rien n'empêche alors de répéter le processus pour cette nouvelle suite.

On dira qu'une suite i est construite de manière **auto-consistante** si

– i_0 est tiré uniformément parmi les n situations
– à chaque étape, i_{s-1} est tiré selon la probabilité $p_J(k \mid i_{s-2} \dots i_0)$

On remarque que la probabilité d'obtenir une suite $i = (i_0 \dots i_{t-1})$ de cette manière vaut

$$p_{ac}(i) = p_J(i_{t-1} \mid i_{t-2} \dots i_0) p_J(i_{t-2} \mid i_{t-3} \dots i_0) \dots p_J(i_1 \mid i_0) p_J(i_0) \tag{18}$$

Or il se trouve que cette probabilité est donnée par la fonction $H \circ m$

$$p_{ac}(i) = H(m(i)) \tag{19}$$

Comme aucune hypothèse supplémentaire est faite, on nommera la probabilité p_{ac} **a priori (non informatif) auto-consistant** sur les suites i de longueur donnée t. L'échantillonnage selon cet a priori sera utile plus loin pour l'estimation en face de situations floues.

Remarque Cette manière de générer une suite aléatoire de nombres donne une réponse possible à la question : *Soit une suite donnée, quelle est sa probabilité d'apparition ?* On peut voir facilement que les suites de probabilité auto-consistante maximale sont aussi celles à minimu d'entropie.

On a en passant que pour les suites de longueur t donnée

$$\sum_i H(m(i)) = 1$$

Depuis les relations (18) et (19), on a aussi

$$H(m(i) + e_k) = p_J(k \mid i) H(m(i)) \tag{20}$$

$$H(m(i) + e_k + e_l) = p_J(k \mid i) p_J(l \mid (i,k)) H(m(i)) \tag{21}$$

5 Généralisation aux situations floues

Dans la réalité, les situations sont souvent mesurées par des instruments bruités. Au lieu de connaître la situation à une étape s, l'observateur est seulement en présence d'un résultat de mesure x provenant d'un modèle de mesure $\mathrm{p}(x \mid k)$. Pour x donné, ce vecteur, dont les indices sont les situations, se nomme souvent fonction de vraisemblance. Pour une suite d'étapes $s = 0 \dots t-1$, les observations[2] $(x_0 \dots x_{t-1})$ donnent une **matrice de vraisemblance** dont le temps est l'indice ligne et la situation l'indice colonne

$$\ell^s(k) = \mathrm{p}(x_s \mid k) \tag{22}$$

Reprenons la situation sur l'hypersphère. Supposons qu'au temps s, la probabilité a priori sur $\mathbb{S}^{n-1}$ soit $d\eta_s(\psi)$. L'apparition d'un résultat x_s pour le modèle de mesure $\mathrm{p}(x \mid k)$, composé avec le modèle canonique $\mathrm{p}(k \mid \psi) = \psi_k^2$ donne par le théorème de Bayes, un a posteriori qui sera repris comme a priori à l'étape suivante :

$$d\eta_{s+1}(\psi) = \frac{\sum_k \ell^s(k) \psi_k^2 d\eta_s(\psi)}{\sum_k \ell^s(k) \int \psi_k^2 d\eta_s(\psi)} \tag{23}$$

[2] Le modèle de mesure peut varier à chaque étape.

On peut donc écrire

$$d\eta_t(\psi) = \frac{\sum_i \ell^0(i_0)\dots\ell^{t-1}(i_{t-1})\psi^2_{i_0}\dots\psi^2_{i_{t-1}}d\mu(\psi)}{\sum_i \ell^0(i_0)\dots\ell^{t-1}(i_{t-1})\int\psi^2_{i_0}\dots\psi^2_{i_{t-1}}d\mu(\psi)} \tag{24}$$

Posons

$$\ell(i) = \prod_s \ell^s(i_s)$$

et utilisons la définition de la fonction H pour obtenir les estimateurs à $t-1$

$$\mathbb{E}(p_k) = \frac{\sum_i \ell(i)H(m(i)+e_k)}{\sum_i \ell(i)H(m(i))} \tag{25}$$

$$\mathbb{E}(p_k p_l) = \frac{\sum_i \ell(i)H(m(i)+e_k+e_l)}{\sum_i \ell(i)H(m(i))} \tag{26}$$

qui s'écrivent aussi

$$\mathbb{E}(p_k) = \frac{\sum_i p_J(k \mid i)\ell(i)H(m(i))}{\sum_i \ell(i)H(m(i))} \tag{27}$$

$$\mathbb{E}(p_k p_l) = \frac{\sum_i p_J(k \mid i)p_J(l \mid (i,k))\ell(i)H(m(i))}{\sum_i \ell(i)H(m(i))} \tag{28}$$

Les équations précedentes laissent subodorer une nouvelle utilisation du théorème de Bayes. En effet, depuis l'a priori auto-consistant p_{ac} sur les suites de longueur t, et le modèle de mesure produit donnant la fonction de vraisemblance ℓ, on peut former l'a posteriori sur les suites

$$p_{ac}(i \mid \ell) = \frac{\ell(i)p_{ac}(i)}{\sum_j \ell(j)p_{ac}(j)} \tag{29}$$

et il vient

$$\mathbb{E}(p_k) = \sum_i p_J(k \mid i)p_{ac}(i \mid \ell) \tag{30}$$

$$\mathbb{E}(p_k p_l) = \sum_i p_J(k \mid i)p_J(l \mid (i,k))p_{ac}(i \mid \ell) \tag{31}$$

Les équations (30) et (31) sont des intégrales qui peuvent être calculées par Monte Carlo. Il est possible d'échantillonner l'a posteriori $p_{ac}(i \mid \ell)$ de la manière suivante :

- tirer i_0 selon les poids $\ell^0(k)$
- tirer i_1 selon les poids $\ell^1(k)p_J(k \mid i_0)$
- ...
- tirer i_{t-1} selon les poids $\ell^{t-1}(k)p_J(k \mid i_{t-2}\dots i_0)$

Un échantillonnage de r suites i_s de ce type donne les estimateurs MC

$$\widehat{\mathbb{E}(p_k)} = \frac{1}{r}\sum_s p_J(k \mid i_s) \tag{32}$$

$$\widehat{\mathbb{E}(p_k p_l)} = \frac{1}{r}\sum_s p_J(k \mid i_s)p_J(l \mid (i_s,k)) \tag{33}$$

ainsi que des incertitudes MC

$$u_{\mathbb{E}(p_k)} = \frac{1}{\sqrt{r}}\sqrt{\frac{1}{r}\sum_s p_J(k \mid i_s)^2 - \widehat{\mathbb{E}(p_k)}^2} \tag{34}$$

$$u_{\mathbb{E}(p_k p_l)} = \frac{1}{\sqrt{r}}\sqrt{\frac{1}{r}\sum_s p_J(k \mid i_s)^2 p_J(l \mid (i_s, l))^2 - \widehat{\mathbb{E}(p_k p_l)}^2} \tag{35}$$

Remarque La simulation de cet échantillonnage est très facile par exemple en C^{++}, et donne d'excellents résultats pour des matrices de vraisemblance avec 3 situations et 10000 temps d'obsrervation avec un r de l'ordre du million. L'intérêt de l'échantillonnage sur l'a posteriori est de localiser directement les suites contribuant beaucoup à la somme générale. Rappelons simplement que les sommes (30) et (31) contiennent n^t termes.

6 Conclusion

Le remplacement de p_k par $\psi_k = \sqrt{p_k}$ rend l'a priori de Jeffreys uniforme. D'une manière générale, les grandeurs de type probabilistes admettent une paramétrisation par la racine carrée qui est leur expression naturelle. On retrouve ce phénomène en mécanique quantique, où la probabilité s'exprime comme carré du module d'une fonction d'onde, d'où le choix de la lettre grecque ψ pour les points de l'hypersphère.

La généralisation des fonctions bêta donnée par la fonction H permet de jeter un pont entre le calcul continu sur les fonctions d'onde et les sommes sur les séquences, qui sont calculables par MC.

Une utilisation de ce travail est en développement pour les processus de Markov stationnaires.

7 Résumé

7.1 Vu par $\psi \in \mathbb{S}^{n-1}$

$$d\eta^0(\psi) = d\mu(\psi) \text{ invariante} \tag{36}$$

$$d\eta^t(\psi) = d\eta^{t-1}(\psi \mid \ell^{t-1}) = \frac{\sum_k \ell^{t-1}(k)\psi_k^2 d\eta^{t-1}(\psi)}{\sum_k \ell^{t-1}(k)\int \psi_k^2 d\eta^{t-1}(\psi)} \tag{37}$$

$$\mathbb{E}^t(p_k \mid \ell) = \int \psi_k^2 d\eta^t(\psi) \tag{38}$$

$$\mathbb{E}^t(p_k p_l \mid \ell) = \int \psi_k^2 \psi_l^2 d\eta^t(\psi) \tag{39}$$

7.2 Vu par $i \in \{0 \dots n-1\}^{\mathbb{N}}$

$$
\begin{aligned}
i^t &= (i_0 \dots i_{t-1}) && (40)\\
p^t_{ac}(i) &= p_J(i_{t-1} \mid i^{t-1}) p^{t-1}_{ac}(i) && (41)\\
\ell^t(i) &= \prod_{s=0}^{t-1} \ell^s(i_s) && (42)\\
p^t_{ac}(i \mid \ell) &= \frac{\ell^t(i) p^t_{ac}(i)}{\sum_{i^t} \ell^t(i) p^t_{ac}(i)} && (43)\\
\mathbb{E}^t(p_k \mid \ell) &= \sum_{i^t} p_J(k \mid i^t) p^t_{ac}(i \mid \ell) && (44)\\
\mathbb{E}^t(p_k p_l \mid \ell) &= \sum_{i^t} p_J(l \mid i^t k) p_J(k \mid i^t) p^t_{ac}(i \mid \ell) && (45)
\end{aligned}
$$