

Chercher une aiguille dans une botte de foin

Bruno Jedynak (Bruno.Jedynak@univ-lille1.fr)

November 6, 2001

1 De quoi s'agit-il ?

Je me propose d'étudier le modèle suivant qui provient de [1]. C'est une version simplifiée d'une modélisation issue d'un problème d'identification d'une courbe dans une image. Soient donc

$$\begin{pmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,n} \\ X_{2,1} & X_{2,2} & \dots & X_{2,n} \\ \vdots & \vdots & & \vdots \\ X_{m+1,1} & X_{m+1,2} & \dots & X_{m+1,n} \end{pmatrix}$$

$n(m+1)$, avec $n > 1$ et $m > 1$ variables aléatoires indépendantes (v.a.i) définies sur un même espace $(\Omega, \mathcal{F}, \mu)$. La loi jointe de ces v.a.i est un produit dont les marginales sont définies comme suit: sur chacune des $m+1$ lignes, les v.a.i sont de même loi; il y a une ligne particulière, que je note l , $1 \leq l \leq m+1$ pour laquelle les n v.a.i $X_{l,1}, \dots, X_{l,n}$ sont de loi $f_1 d\mu$; toutes les autres v.a.i sont de loi $f_0 d\mu$. f_0, f_1 sont deux densités.

La question statistique qui m'intéresse est l'estimation du paramètre l au vu d'une réalisation des $n \times (m+1)$ v.a.i définies ci-dessus.

La figure 1 permet de visualiser le problème. L'image de gauche est formée de $m+1 = 30 \times 30 = 900$ imagettes. Chacune des imagettes est composée $n = 10 \times 10 = 100$ pixels. Le niveau de gris de chaque pixel est la réalisation d'une v.a. Toutes les v.a. sont indépendantes entre elles. Pour une des imagettes, appelée la cible, visiblement plus claire, située dans la partie basse, au centre de l'image, les simulations proviennent d'une loi Normale notée f_1 , de moyenne $m_1 = 140$ et d'écart type $\sigma_1 = 20$. Pour tous les autres pixels, les simulations proviennent d'une loi Normale notée f_0 de moyenne $m_0 = 100$ et d'écart type $\sigma_0 = 20$. Le problème décrit plus haut

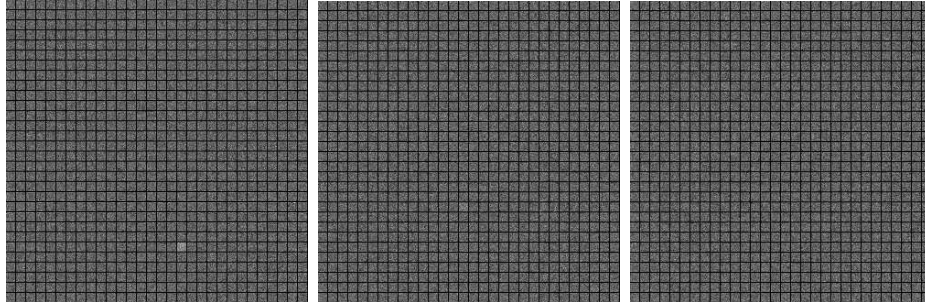


Figure 1:

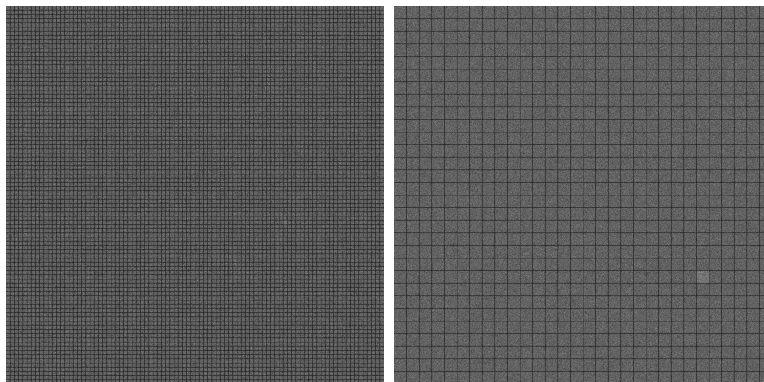


Figure 2:

est celui de l'identification de l'unique image constituée de simulations de v.a. de loi f_1 . L'identification visuelle (en disposant d'une reproduction suffisamment bonne) est facile. L'image du centre est générée avec les mêmes paramètres, à l'exception de m_1 , qui vaut 120. La cible se trouve aussi dans la partie basse, au centre de l'image et son identification est à la limite des capacités visuelles. L'image de droite est simulée de la même manière, mais avec cette fois-ci, $m_1 = 110$. La cible n'est plus identifiable visuellement. Quand les lois se "rapprochent", l'identification de la cible est de plus en plus difficile. Les deux images de la figure 2 sont simulées avec les mêmes densités f_0 et f_1 que pour l'image du centre de la figure 1. A gauche, $n = 10 \times 10$ et $m + 1 = 90 \times 90$, et à droite, $n = 30 \times 30$ et $m + 1 = 30 \times 30$. Pour n fixé, l'augmentation de m , c'est à dire du nombre de distracteurs rend l'identification de la cible de plus en plus difficile. Pour m fixé, quand n augmente, l'identification visuelle de la cible est de plus en plus facile.

Quelles sont les performances maximales d'une méthode d'identification de la cible comme fonction des quatre paramètres du problème qui sont: les nombres m et n et les densités f_0 et f_1 ?

La contribution principale de ce papier est le Théorème 2 qui permet de répondre à la question posée lorsque simultanément m et n tendent vers l'infini. Je présente de plus un résultat non asymptotique: Le théorème 1 qui précise et généralise le Corollaire 3 de [1]. La suite de ce papier est organisée comme suit. Dans la section 2 j'explicité l'estimateur du maximum de vraisemblance. Ce serait l'estimateur obtenu dans un cadre bayésien avec la fonction de perte qui pénalise de la même manière toutes les erreurs d'identification de la cible. Ses performances asymptotiques et non asymptotiques sont présentées dans la section 3. Les preuves sont reportées dans la section 5. Enfin, la section 4 met en regard les résultats obtenus et les simulations.

2 L'estimateur du maximum de vraisemblance (e.m.v)

Il est naturel de regarder la vraisemblance. Celle-ci s'écrit pour une ligne quelconque l , $1 \leq l \leq m + 1$ et une réalisation $\omega \in \Omega$:

$$V(\omega, l) = \prod_{j=1}^n f_1(X_{l,j}(\omega)) \prod_{i=1, i \neq l}^{m+1} \prod_{j=1}^n f_0(X_{i,j}(\omega))$$

L'estimateur du maximum de vraisemblance, noté $\hat{l}(\omega)$, pour \mathbf{l} est alors,

pour chaque $\omega \in \Omega$, une valeur l qui réalise le maximum de $V(\omega, l)$. En cas d'égalité, je choisis (par exemple) la plus petite valeur l qui réalise le maximum de $V(\omega, l)$. Je donne maintenant une écriture un peu plus simple de cet estimateur. Pour que tout soit bien défini, je me place tout d'abord sur

$$A = \{(\omega, l) \in (\Omega, \{1, \dots, m+1\}), V(\omega, l) > 0\}$$

Il vient alors,

$$\begin{aligned} \ln V(\omega, l) &= \sum_{j=1}^n \ln f_1(X_{l,j}(\omega)) + \sum_{i=1, i \neq l}^{m+1} \sum_{j=1}^n \ln f_0(X_{i,j}(\omega)) \\ &= \sum_{j=1}^n \ln f_1(X_{l,j}(\omega)) + \sum_{i=1}^{m+1} \sum_{j=1}^n \ln f_0(X_{i,j}(\omega)) - \sum_{j=1}^n \ln f_0(X_{l,j}(\omega)) \\ &= \sum_{j=1}^n \ln \frac{f_1}{f_0}(X_{l,j}(\omega)) + \sum_{i=1}^{m+1} \sum_{j=1}^n \ln f_0(X_{i,j}(\omega)) \end{aligned}$$

Ainsi sur A , $\hat{l}(\omega)$ vérifie, avec la convention indiquée précédemment en cas d'égalité:

$$\hat{l}(\omega) = \arg \max_{1 \leq l \leq m+1} \sum_{j=1}^n \ln \frac{f_1}{f_0}(X_{l,j}(\omega)) \quad (1)$$

Je peux maintenant prolonger sur le complémentaire de A cette définition en posant

$$\forall x > 0, \ln \frac{x}{0} = +\infty, \ln \frac{0}{x} = -\infty \quad (2)$$

L'estimateur du maximum de vraisemblance (e.m.v) $\hat{l}(\omega)$ est alors défini μ p.s. En effet, f_0 et f_1 ne peuvent pas (p.s) s'annuler simultanément pour une valeur $(X_{i,j}(\omega))$ car la v.a. $X_{i,j}$ est soit de loi $f_0 d\mu$, soit de loi $f_1 d\mu$. Afin de simplifier les notations, j'omet dans la suite la dépendance en ω . Ainsi, je pose,

$$\hat{l} = \arg \max_{1 \leq l \leq m+1} \sum_{j=1}^n \ln \frac{f_1}{f_0}(X_{l,j}) \quad (3)$$

Dans un cadre bayésien, la seule fonction de coût intéressante ici est la fonction binaire qui vaut 0 si la cible est identifiée, et 1 sinon. L'estimateur bayésien optimal coïncide alors avec (3). En ce sens, les performances de l'e.m.v. que nous étudions maintenant sont les meilleures performances accessibles pour estimer la cible.

3 Performances de l'e.m.v

Quelles sont les performances de l'e.m.v.(3)? Et en particulier que se passe t'il lorsque simultanément $m \rightarrow +\infty$ et $n \rightarrow +\infty$? Une bonne mesure de l'erreur commise en estimant \mathbf{l} par $\hat{\mathbf{l}}$ est

$$A_{m,n} = Pr(\hat{\mathbf{l}} \neq \mathbf{l}) \quad (4)$$

$$\geq Pr\left(\max_{1 \leq l \leq m+1, l \neq 1} \sum_{j=1}^n \ln \frac{f_1}{f_0}(X_{l,j}) > \sum_{j=1}^n \ln \frac{f_1}{f_0}(X_{1,j})\right) \quad (5)$$

Remarquons que (4) est le risque bayésien de l'e.m.v, vu comme un estimateur bayésien.

Les quantités (4) et (5) coïncident sauf en cas d'égalité dans le choix de $\hat{\mathbf{l}}$. Asymptotiquement, l'égalité à lieu $\mu p.s.$ J'étudie donc dans la suite la quantité (5) que je renomme $A_{m,n}$.

L'étude asymptotique de la quantité $A_{m,n}$ est l'étude de la convergence en probabilité de l'estimateur $\hat{\mathbf{l}}$ vers la cible \mathbf{l} . Remarquons que l'on ne peut pas parler de convergence presque sûre ici puisque cet estimateur n'est pas défini quand $m \rightarrow +\infty$.

Examinons tout d'abord deux cas spéciaux:

Si f_0 et f_1 sont étrangères, c'est à dire que quelque soient i et j , soit $f_0(X_{i,j}) = 0$, soit $f_1(X_{i,j}) = 0$ μ p.s alors les conventions (2) permettent alors bien d'obtenir que $\forall n \geq 1, \forall m \geq 1, A_{m,n} = 0$.

Si maintenant $f_0 = f_1, \mu$ p.s. alors

$$\forall n \geq 1, A_{m,n} = \frac{m}{m+1} \rightarrow 1 \text{ lorsque } m \rightarrow \infty$$

Il est clairement impossible d'estimer correctement \mathbf{l} .

Notons

$$Z_i = \frac{1}{n} \sum_{j=1}^n \ln \frac{f_1}{f_0}(X_{i,j})$$

on a alors

$$A_{m,n} = Pr\left(\max_{1 \leq l \leq m+1, l \neq 1} Z_l > Z_1\right)$$

Nous pouvons maintenant présenter un premier résultat qui exprime que si, asymptotiquement, m ne croit pas trop vite vers l'infini avec n alors l'e.m.v est consistant.

Théorème 1

$$\forall n \geq 1, \forall m \geq 1, A_{m,n} \leq mg(f_0, f_1)^n$$

avec

$$g(f_0, f_1) = \left(\int \sqrt{f_0 f_1} d\mu \right)^2$$

ainsi,

si $\lim_{m \rightarrow \infty, n \rightarrow \infty} mg(f_1, f_0)^n = 0$ alors l'e.m.v est consistant.

Remarquons que $g(f_1, f_0) = (1 - \text{Hellinger}(f_0, f_1))^2$. Ainsi, plus la distance de Hellinger entre les densités f_0 et f_1 est grande, et moins l'e.m.v est sensible à l'augmentation du nombre des distracteurs. Ce qui est conforme à l'intuition.

Présentons maintenant un résultat purement asymptotique.

Théorème 2 *Le comportement asymptotique de la quantité $A_{m,n}$ est déterminé de la manière suivante:*

Soit $K(f_1, f_0)$ la distance de Kulback-Leibler entre f_1 et f_0 c'est-à-dire

$$\begin{aligned} K(f_1, f_0) &= \int f_1 \log \frac{f_1}{f_0} d\mu \\ &\quad \text{si le support de } f_1 \text{ est inclus dans celui de } f_0 \text{ } \mu \text{ p.s.} \\ &= +\infty \text{ sinon} \end{aligned}$$

1. si $K(f_1, f_0) = +\infty$ alors

$$\lim_{m \rightarrow \infty, n \rightarrow \infty} A_{m,n} = 0$$

2. si $K(f_1, f_0) = 0$ alors

$$\lim_{m \rightarrow \infty, n \rightarrow \infty} A_{m,n} = 1$$

3. si $0 < K(f_1, f_0) < \infty$ alors

(a) si

$$\exists \epsilon > 0, \lim_{m \rightarrow \infty, n \rightarrow \infty} m e^{-n(K(f_1, f_0) + \epsilon)} = 0$$

alors

$$\lim_{m \rightarrow \infty, n \rightarrow \infty} A_{m,n} = 0$$

(b) si

$$\exists \epsilon > 0, \lim_{m \rightarrow \infty, n \rightarrow \infty} m e^{-n(K(f_1, f_0) - \epsilon)} = \infty$$

alors

$$\lim_{m \rightarrow \infty, n \rightarrow \infty} A_{m,n} = 1$$

Ainsi, en dehors des cas extrêmes 1. et 2., le comportement asymptotique de l'erreur d'estimation est dicté par les vitesses relatives de croissance vers l'infini de m et de n .

4 Quelques simulations

Que nous disent les résultats énoncés pour le problème des figures 1 et 2 ? L'application numérique du théorème 1 permet d'obtenir les inégalités ci-dessous:

Figure 1 à gauche	$Pr(\hat{l} \neq 1) \leq 3 \times 10^{-41}$
Figure 1 au centre	$Pr(\hat{l} \neq 1) \leq 2 \times 10^{-8}$
Figure 1 à droite	$Pr(\hat{l} \neq 1) \leq 1.7$
Figure 2 à gauche	$Pr(\hat{l} \neq 1) \leq 2 \times 10^{-7}$
Figure 2 à droite	$Pr(\hat{l} \neq 1) \leq 10^{-95}$

Ainsi, dans tous les cas, sauf pour l'image de la figure 1 à droite, le théorème 1 garantit que les performances de l'estimateur du maximum de vraisemblance sont excellentes. Dans le cas indiqué, le théorème 1 est inutile.

Nous présentons maintenant des simulations qui permettent de comprendre comment utiliser le théorème 2 pour des grandes valeurs de m et n .

La figure 3 présente pour 4 choix du couple (f_0, f_1) , un choix par graphique, l'évolution de la quantité $A_{m,n}$ en fonction de n . Le nombre de distracteurs est fixé à $m = 1000$. La ligne brisée verticale indique la valeur n^* de n pour laquelle $m \exp(-n^* K(f_1, f_0)) = 1$. Nous remarquons dans chacun des 4 cas présentés que la valeur n^* indique une inflexion dans les performances de l'e.m.v. L'estimateur devient performant pour $n > n^*$. La figure 4 donne les valeurs numériques pour les deux graphiques du haut.

Nous remarquons sur ces deux tableaux que la majoration donnée par le théorème 1 (troisième colonne) est beaucoup trop grossière pour les petites valeurs de n . Le résultat est intéressant (car ≤ 1 !) dans le premier cas à partir de $n = 512$ et dans le second à partir de $n = 128$. Nous observons sur la dernière colonne de ces tableaux que le comportement de

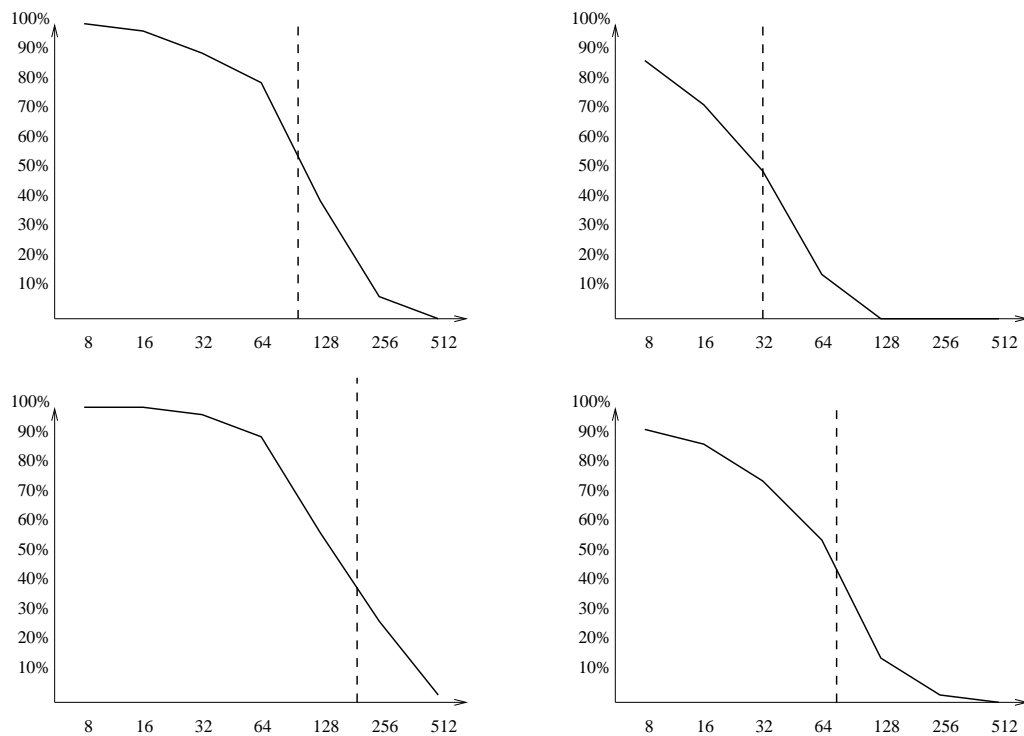


Figure 3: Simulation de l'erreur de l'estimateur du maximum de vraisemblance. Le nombre de valeurs pour chaque ligne (paramètre n) figure en abscisse. Le nombre de distracteurs (paramètre m) est fixé à 1000. L'axe des ordonnées représente les valeurs possibles de la probabilité de l'erreur d'estimation $A_{1000,n}$. La ligne en pointillés verticale indique la valeur de l'abscisse n pour laquelle la quantité $1000 \exp(-nK(f_1, f_0)) = 1$. **En haut à gauche** les lois f_0 et f_1 sont des lois de Bernouilli de paramètres $p_0 = 0.1$ et $p_1 = 0.2$. **En haut à droite** les lois f_0 et f_1 sont des lois de Bernouilli de paramètres $p_0 = 0.1$ et $p_1 = 0.3$. **En bas à gauche** les lois f_0 et f_1 sont des lois Normales de même variance $\sigma^2 = 1$ et dont les moyennes valent respectivement 0 et 0.25. $K(f_1, f_0) = 0.03$. **En bas à droite** les lois f_0 et f_1 sont des lois Normales de même moyenne $m = 0$ et dont les variances valent respectivement 1 et 1.69.

n	$A_{m,n}$	mg^n	$m \exp(-nK(f_1, f_0))$
16	0.98	723	491
32	0.90	523	241
64	0.80	274	58
128	0.40	75	3.4
256	0.08	5.6	0.01
512	0.00	0.03	10^{-7}

n	$A_{m,n}$	mg^n	$m \exp(-nK(f_1, f_0))$
8	0.98	583	292
16	0.72	360	85
32	0.50	116	7
64	0.14	13	0.05
128	0.00	0.18	3×10^{-6}
256	0.00	310^{-5}	8×10^{-15}

Figure 4: Valeurs numériques de la simulation de la probabilité d'erreur de l'estimateur du maximum de vraisemblance. **En haut:** f_0 et f_1 sont des lois de Bernoulli. Leurs paramètres sont respectivement $p_0 = 0.1$ et $p_1 = 0.2$. Ainsi, $g = 0.98$ et $K(f_1, f_0) = 0.044$. $m = 1000$ est fixé, n prend successivement les valeurs 16,32,64,128,256 et 512. $A_{m,n}$ est estimé à chaque fois en effectuant 100 simulations et en effectuant une moyenne empirique. **En bas:** f_0 et f_1 sont aussi des lois de Bernoulli mais leurs paramètres sont respectivement $p_0 = 0.1$ et $p_1 = 0.3$. Ainsi, $g = 0.93$ et $K(f_1, f_0) = 0.154$. $m = 1000$ est fixé, n prend successivement les valeurs 16,32,64,128,256 et 512. $A_{m,n}$ est estimé à chaque fois en effectuant 100 simulations et en effectuant une moyenne empirique.

l'erreur d'estimation $A_{m,n}$ est corrélée à l'équilibre entre les quantités m et $\exp(nK(f_1, f_0))$. L'erreur est importante quand la première dépasse la seconde, elle est faible dans le cas contraire. Ceci est tout à fait cohérent avec le résultat asymptotique du théorème 2 comme nous l'avons indiqué précédemment.

5 Preuve des résultats

Preuve du théorème 1: La preuve de ce résultat que nous présentons maintenant utilise les majorations élémentaires ci-dessous ainsi que des techniques classique en théorie des grandes déviations.

$$A_{m,n} = Pr(\cup_{1 \leq l \leq m+1, l \neq 1} \{Z_l > Z_1\}) \quad (6)$$

$$\leq \sum_{1 \leq l \leq m+1, l \neq 1} Pr(Z_l > Z_1) \quad (7)$$

$$\leq m Pr(Z_l > Z_1) \text{ avec } l \neq 1 \quad (8)$$

(6) est une réécriture de la définition de $A_{m,n}$, (7) est obtenu en majorant la probabilité d'une union par la somme des probabilités et (8) car toutes les v.a.i Z_l pour $l \neq 1$ sont de même loi. Il reste à montrer que

$$Pr(Z_l > Z_1) \leq g(f_0, f_1)^n \text{ pour } l \neq 1 \quad (9)$$

$$Pr(Z_l > Z_1) = Pr\left(\sum_{j=1}^n \ln \frac{f_1}{f_0}(Y_{lj}) > \sum_{j=1}^n \ln \frac{f_1}{f_0}(Y_{1j})\right) \quad (10)$$

$$= Pr\left(\exp t\left(\sum_{j=1}^n \ln \frac{f_1}{f_0}(Y_{lj}) - \ln \frac{f_1}{f_0}(Y_{1j})\right) > 1\right), \forall t > 0 \quad (11)$$

$$\leq E\left(\exp\left(t \sum_{j=1}^n \left(\ln \frac{f_1}{f_0}(Y_{lj}) - \ln \frac{f_1}{f_0}(Y_{1j})\right)\right)\right), \forall t > 0 \quad (12)$$

$$\leq \left(E\left(\exp\left(t \ln \frac{f_1}{f_0}(Y_0)\right)\right)\right) E\left(\exp\left(-t \ln \frac{f_1}{f_0}(Y_1)\right)\right)^n, \forall t > 0 \quad (13)$$

$$Y_0 \text{ et } Y_1 \text{ sont des v.a.i de loi resp. } f_0 d\mu \text{ et } f_1 d\mu \quad (14)$$

$$\leq \left(E\left(\exp\left(t \ln \frac{f_1}{f_0}(Y_0)\right)\right)\right) E\left(\exp\left((1-t) \ln \frac{f_1}{f_0}(Y'_0)\right)\right)^n, \forall t > 0 \quad (15)$$

$$Y'_0 \text{ est une v.a. indépendante de } Y_0 \text{ et de loi } f_0 d\mu. \quad (16)$$

$$\leq \left(E\left(\exp\left(\frac{1}{2} \ln \frac{f_1}{f_0}(Y_0)\right)\right)\right)^{2n} = \left(\int \sqrt{f_0 f_1} d\mu\right)^{2n} \quad (17)$$

Rappelons que dans (10), toutes les variables aléatoires Y_l sont indépendantes entre elles. Les Y_l , pour $l \neq 1$ sont de loi $f_0 d\mu$ et les Y_1 sont de loi $f_1 d\mu$. (12) est l'inégalité de Chebyshev. (13) car les v.a. sont indépendantes. (15) car

$$E(\exp(-t \ln \frac{f_1}{f_0}(Y_1))) = E(\exp((1-t) \ln \frac{f_1}{f_0}(Y'_0)))$$

(17) est obtenu en minimisant (15) par rapport à t . Le minimum unique de (15) en t est obtenu pour la valeur $t = \frac{1}{2}$. Ce qui conclut la preuve du théorème 1.

Preuve du Théorème 2:

La situation 2. est équivalente à $f_0 = f_1, \mu p.s.$ Dans ce cas

$$\forall n \geq 1, A_{m,n} = \frac{m}{m+1} \rightarrow 1 \text{ lorsque } m \rightarrow \infty \quad (18)$$

Examinons maintenant le cas 1. Rappelons tout d'abord que

$$Z_1 = \sum_{j=1}^n \ln \frac{f_1}{f_0}(X_{1,j}) \quad (19)$$

où les v.a. $X_{1,j}$ sont indépendantes entre elles et de loi $f_1 d\mu$. Si $K(f_1, f_0) = +\infty$ alors pour n suffisamment grand, on a $\mu p.s.$, $Z_l = \infty$ et par suite

$$A_{m,n} = Pr(\max_{1 \leq l \leq m+1, l \neq 1} Z_l > Z_1) = 0$$

Supposons maintenant que $0 < K(f_1, f_0) < +\infty$. Nous allons procéder par étape pour trouver un équivalent asymptotique de la quantité $A_{m,n}$. L'argument principal est un résultat de grandes déviation précis ([2]), c'est à dire fournissant un équivalent asymptotique.

La loi des grands nombres permet d'affirmer que la quantité Z_1 converge en probabilité (ce qui nous suffit) lorsque $n \rightarrow +\infty$ vers son espérance qui est $K(f_1, f_0)$. Ainsi, pour chaque m , il existe deux suites notées $u_{m,1}(n)$ et $u_{m,2}(n)$ qui tendent vers 0 lorsque $n \rightarrow \infty$ et qui vérifient

$$1 - A_{m,n} = Pr(\max_{1 \leq l \leq m+1, l \neq 1} Z_l < Z_1) \quad (20)$$

$$= Pr(\max_{1 \leq l \leq m+1, l \neq 1} Z_l < K(f_1, f_0) + u_{m,1}(n)) + u_{m,2}(n) \quad (21)$$

La probabilité dans (21) se réécrit, de par l'hypothèse d'indépendance.

$$Pr(\max_{1 \leq l \leq m+1, l \neq 1} Z_l < K(f_1, f_0) + u_{m,1}(n)) \quad (22)$$

$$= (1 - Pr(Z_l > K(f_1, f_0) + u_{m,1}(n)))^m \text{ avec } l \neq 1 \quad (23)$$

Examinons maintenant le comportement asymptotique, quand $n \rightarrow \infty$ de la quantité

$$Pr(Z_l > K(f_1, f_0) + u_{m,1}(n)) \text{ pour } l \neq 1 \quad (24)$$

D'une part, Z_l , est une moyenne empirique de variables i.i.d. d'espérance $-K(f_0, f_1)$. Cette quantité est strictement négative car $f_0 \neq f_1$. D'autre part, pour chaque m et pour n suffisamment grand, $K(f_1, f_0) + u_{m,1}(n) > 0$. En effet, $K(f_1, f_0) > 0$ car $f_0 \neq f_1$ et $\lim_{n \rightarrow \infty} u_{m,1}(n) = 0$.

Ainsi, d'après la loi (faible suffit) des grands nombres

$$\lim_{n \rightarrow \infty} Pr(Z_l > K(f_1, f_0) + u_{m,1}(n)) = 0$$

La vitesse de convergence est donné par le théorème des grandes déviations. En appliquant le théorème 1 de ([2]), nous obtenons que l'on peut trouver des suites $u_{m,3}(n)$ et $u_{m,4}(n)$ qui tendent vers 0 lorsque $n \rightarrow \infty$ et telles que

$$Pr(Z_l > K(f_1, f_0) + u_{m,1}(n)) = \quad (25)$$

$$\exp(-n(h(K(f_1, f_0) + u_{m,1}(n)) + u_{m,3}(n))) \frac{1}{\sqrt{2\pi n}} (1 + u_{m,4}(n)) \quad (26)$$

$$\text{avec } h(a) = \sup_{u>0} [ua - \int \exp(u \ln \frac{f_1}{f_0}) f_0 d\mu] \quad (27)$$

La fonction h est une transformée de Cramer. Elle est strictement convexe donc en particulier continue et on vérifie en dérivant que

$$h(K(f_1, f_0)) = K(f_1, f_0) \quad (28)$$

Ainsi, pour chaque m , on peut trouver une nouvelle suite $u_{m,5}(n)$ qui tend vers 0 lorsque $n \rightarrow \infty$ et telle que:

$$\begin{aligned} h(K(f_1, f_0) + u_{m,1}(n)) &= h(K(f_1, f_0)) + u_{m,5}(n) \text{ par continuité} \\ &= K(f_1, f_0) + u_{m,5}(n) \text{ d'après (28)} \end{aligned}$$

Notons, pour chaque $m \geq 1$,

$$t(m, n) = \exp(-n(K(f_1, f_0) + u_{m,5}(n)) + u_{m,3}(n)) \frac{1}{\sqrt{2\pi n}} (1 + u_{m,4}(n))$$

Nous avons donc obtenu, pour chaque $m \geq 1$,

$$1 - A_{m,n} = (1 - t(m, n))^m + u_{m,2}(n)$$

Or, pour chaque m , $t(m, n)$ tend vers 0 lorsque $n \rightarrow \infty$, il existe donc une nouvelle suite $u_{m,6}(n)$ qui tend vers 0 lorsque $n \rightarrow \infty$ et telle que

$$1 - A_{m,n} = \exp(-mt(m, n)(1 + u_{m,6}(n)) + u_{m,2}(n))$$

Ainsi, le comportement asymptotique de $A_{m,n}$ est-il dicté par le comportement asymptotique du terme $mt(m, n)$, ce qui permet de conclure.

References

- [1] James M. Coughlan Alan L. Yuille. Fundamental limits of bayesian inference: Order parameters and phase transitions for road tracking. *IEEE Transactions on PAMI*, 22(2):160–173, Février 2000.
- [2] R.R. Bahadur and R. Ranga Rao. On deviations of the sample mean. *Ann. Math. Stat.*, 31:1015–1027, 1960.