

Approche bayésienne empirique dans les tables de contingence de dimension deux

Melle Nadja EL SAADI • Institut National de la Planification et de la statistique.

RESUME

Une analyse bayésienne est appliquée aux paramètres du modèle loglinéaire saturé. Les moyennes a priori sont choisies de façon à être les termes d'association d'un modèle simple pour les données ordinales: le modèle d'association $L \times L$ proposé par GOODMAN (1979). Des méthodes empiriques pour estimer les paramètres des lois a priori sont discutées puis comparées à travers des exemples illustratifs.

MOTS-CLES : modèle loglinéaire - variable ordinale - modèle $L \times L$ - loi a priori - loi a posteriori - estimateur de Bayes - Algorithme EM.

I- INTRODUCTION

Laird (1978) proposa d'estimer les proportions d'une table de contingence de dimension 2 à travers une analyse bayésienne appliquée aux paramètres du modèle loglinéaire saturé:

$$\text{Log } m_{ij} = \mu + \lambda_i^x + \lambda_j^y + \lambda_{ij}$$

Laird adopta cette approche en présumant que les paramètres marginaux $\{\lambda_i^x\}$ et $\{\lambda_j^y\}$ ont des lois a priori uniformes (impropres) indépendantes tout en étant soumis aux contraintes :

$$\sum_i \lambda_i^x = \sum_j \lambda_j^y = 0$$

Elle supposa aussi que les paramètres d'association $\{\lambda_{ij}\}$ sont indépendants et identiquement distribués suivant des lois $N(0, \sigma^2)$. Dans notre travail, nous présentons une adaptation de la méthode de Laird dans laquelle les relations ordinales sont mises en relief. Pour ce faire, nous supposons que les moyennes a priori pour les $\{\lambda_{ij}\}$ sont les termes d'association d'un modèle ordinal simple : le modèle $L \times L$. Soit $n' = (n_{11}, \dots, n_{rc})$ l'échantillon multinomial des observations, obtenu à partir de la classification croisée de deux variables ordinales X

et Y pour une table $r \times c$.

$n' = (n_{11}, \dots, n_{rc})$ est la réalisation du vecteur aléatoire

$(N_{11}, \dots, N_{rc})$ où $(N_{11}, \dots, N_{rc}) \sim \text{Mult}(N, \{\pi_{ij}\})$ avec

$N = \sum \sum n_{ij}$ la taille de l'échantillon et π_{ij} la probabilité d'être classé dans la case (i, j) . Soit $m_{ij} = E(N_{ij}) = N \pi_{ij}$ la fréquence espérée.

Le but de notre étude est l'estimation des $\{\pi_{ij}\}$ en utilisant le modèle loglinéaire saturé :

$$\text{Log } m_{ij} = \mu + \lambda_i^x + \lambda_j^y + \lambda_{ij}$$

Pour cela, on suppose toujours que les termes marginaux $\{\lambda_i^x\}$ et $\{\lambda_j^y\}$ sont indépendants et possèdent des lois a priori impropres sur $\mathbf{R}$ tout en satisfaisant :

$$\sum_i \lambda_i^x = \sum_j \lambda_j^y = 0$$

D'autre part, on suppose que les $\{\lambda_{ij}\}$ sont indépendamment et identiquement distribués suivant des lois normales $N(\beta u_i v_j, \sigma^2)$ où les $\{u_i\}$ sont les scores pour les catégories lignes et les $\{v_j\}$ sont les scores pour les catégories colonnes.

Les $\{\lambda_i^x\}$, $\{\lambda_j^y\}$ et $\{\lambda_{ij}\}$ sont supposés mutuellement indépendants.

II - ESTIMATION DES PROPORTIONS DANS LES TABLES DE CONTINGENCE DE DIMENSION 2

Soit $\theta = (\lambda_1^x, \dots, \lambda_{r-1}^x, \lambda_1^y, \dots, \lambda_{c-1}^y, \lambda_{11}, \dots, \lambda_{rc})$ le vecteur des paramètres, la densité a priori $g(\theta)$ est définie par:

$$g(\theta) \propto \exp \left\{ - \sum_i \sum_j (\lambda_{ij} - \beta u_i v_j)^2 / 2\sigma^2 \right\}$$

La fonction de vraisemblance $l(\theta / n)$ s'écrit:

$$l(\theta / n) = \frac{N!}{\prod_{i,j} n_{ij}! N^{n_{ij}}} \exp \left\{ \sum_i \sum_j n_{ij} \log m_{ij} \right\}$$

D'après le théorème de Bayes, on obtient la densité a posteriori :

$$h(\theta / n; \beta, \sigma) \propto \exp \left\{ \sum_i \sum_j n_{ij} \log m_{ij} - \sum_i \sum_j (\lambda_{ij} - \beta u_i v_j)^2 / 2\sigma^2 \right\}$$

THEOREME

L'estimateur de BAYES de θ est θ^* le mode de loi a posteriori $h(\theta / n; \beta, \sigma)$.

Il est solution des équations de vraisemblance suivantes :

$$m_{i+}^* = n_{i+} \quad 1 \leq i \leq r$$

$$m_{+j}^* = n_{+j} \quad 1 \leq j \leq c$$

$$m_{ij}^* = n_{ij} - (\lambda_{ij}^* - \beta u_i v_j) / \sigma^2 \quad 1 \leq i \leq r \\ 1 \leq j \leq c$$

Les $\{m_{ij}^*\}$ satisfaisant ces équations sont uniques. Pour β et σ^2 fixés, la résolution de ces équations se fait numériquement par la méthode de NEWTON.

Connaissant les valeurs de $\theta^* = (\lambda_1^*, \dots, \lambda_{r-1}^*, \lambda_1^{r*}, \dots, \lambda_{c-1}^*, \lambda_{11}^*, \dots, \lambda_{rc}^*)$, on détermine $\pi^* = (\pi_{11}^*, \dots, \pi_{rc}^*)$ l'estimateur de π pour lequel $m_{ij}^* = N\pi_{ij}^*$.

III- METHODES BAYESIENNES EMPIRIQUES POUR L'ESTIMATION DES PARAMETRES DES LOIS A PRIORI

Dans la section précédente, β et σ furent considérés comme fixés. Dans cette section, on va discuter leur problème d'estimation. Pour cela, nous adoptons l'approche suggérée par LAIRD (1978) et qui fut aussi utilisée par CHUANG (1982) pour un modèle différent. Cette approche utilise une méthode empirique de BAYES qui s'appuie uniquement sur les observations.

La densité a posteriori $h(\theta / n; \beta, \sigma)$ est reliée à la densité a priori $g(\theta; \beta, \sigma)$ et à la vraisemblance $l(\theta / n)$ par la relation :

$$h(\theta / n; \beta, \sigma) = \frac{g(\theta; \beta, \sigma) l(\theta / n)}{m(n; \beta, \sigma)}$$

où $m(n; \beta, \sigma)$ est la fonction de probabilité marginale de n .

En considérant la distribution de n comme fonction de β et σ pour un n donné, nous estimons β et σ dans la loi a priori par les valeurs qui maximisent cette vraisemblance marginale.

Pour cela nous utilisons l'algorithme EM et la notion de données incomplètes (Voir DEMPSTER, LAIRD et RUBIN (1977)).

a) Application de l'algorithme EM au problème d'estimation de (β, σ^2) .

Les observations $n = (n_{11}, \dots, n_{rc})$ sont considérées comme données incomplètes, tandis que $\theta = (\lambda_1^x, \dots, \lambda_{r-1}^x, \lambda_1^y, \dots, \lambda_{c-1}^y, \lambda_{11}, \dots, \lambda_{rc})$ sont vus comme données manquantes.

Les données complètes seront donc :

$$x = (n, \theta) = (n_{11}, \dots, n_{rc}, \lambda_1^x, \dots, \lambda_{r-1}^x, \lambda_1^y, \dots, \lambda_{c-1}^y, \lambda_{11}, \dots, \lambda_{rc})$$

La densité associée aux données complètes peut s'écrire :

$$f(x / \beta, \sigma) = \frac{b(x)}{a(\varphi)} \exp \left\{ -(1/2\sigma^2) \left[\sum_{ij} \lambda_{ij}^2 - 2\beta \sum_{ij} \lambda_{ij} u_i v_j \right] \right\}$$

$$\text{où: } \varphi = \begin{bmatrix} \beta \\ 1/\sigma^2 \end{bmatrix}$$

$$b(x) = f(n / \theta) (2\pi)^{-rc/2}$$

$$\text{et: } a(\varphi) = \sigma^{-rc} \exp \left\{ -(1/2\sigma^2) \beta^2 \sum_{ij} u_i^2 v_j^2 \right\}$$

On remarque que cette dernière appartient à la famille exponentielle, les statistiques exhaustives pour φ sont :

$$\begin{aligned} t(x) &= \begin{bmatrix} t_1(x) \\ t_2(x) \end{bmatrix} \\ &= \begin{bmatrix} \sum_{ij} u_i v_j \lambda_{ij} \\ \sum_{ij} \lambda_{ij}^2 \end{bmatrix} \end{aligned}$$

En appliquant l'algorithme EM, on aura :

Etape E:

$$t^{(p)} = \begin{bmatrix} t_1^{(p)} \\ t_2^{(p)} \end{bmatrix} = \begin{bmatrix} E\left(\sum_{ij} u_i v_j \lambda_{ij} / n; \beta^{(p)}, \sigma^{(p)}\right) \\ E\left(\sum_{ij} \lambda_{ij}^2 / n; \beta^{(p)}, \sigma^{(p)}\right) \end{bmatrix}$$

Etape M:

$$\phi^{(p+1)} = \begin{bmatrix} \beta^{(p+1)} \\ \sigma^{(p+1)} \end{bmatrix} \text{ vérifie:}$$

$$t^{(p)} = \begin{bmatrix} t_1^{(p)} \\ t_2^{(p)} \end{bmatrix} = \begin{bmatrix} E\left(\sum_{ij} u_i v_j \lambda_{ij} / \beta^{(p+1)}, \sigma^{(p+1)}\right) \\ E\left(\sum_{ij} \lambda_{ij}^2 / \beta^{(p+1)}, \sigma^{(p+1)}\right) \end{bmatrix}$$

L'étape M donne:

$$\beta^{(p+1)} = \frac{t_1^{(p)}}{\sum_{ij} u_i^2 v_j^2}$$

$$\left(\sigma^{(p+1)}\right)^2 = \frac{t_2^{(p)} - \left(\beta^{(p+1)}\right)^2 \sum_{ij} u_i^2 v_j^2}{rc}$$

Le calcul des espérances conditionnelles

$$\text{et } \begin{cases} t_1^{(p)} = E\left(\sum_{ij} u_i v_j \lambda_{ij} / n; \beta^{(p)}, \sigma^{(p)}\right) \\ t_2^{(p)} = E\left(\sum_{ij} \lambda_{ij}^2 / n; \beta^{(p)}, \sigma^{(p)}\right) \end{cases}$$

nécessite une intégration par rapport à la loi a posteriori mais malheureusement cette dernière ne possède pas une forme simple. Donc une approximation est inévitablement nécessaire : on approche la loi a posteriori par une normale de même mode et dont le logarithme présente la même courbure en ce mode que le logarithme de la vraie loi a posteriori. Dans ce cas, l'évaluation de la constante de normalisation n'est pas nécessaire.

On suppose donc que sachant n , θ suit une $N(\theta^*, \Sigma^*)$ où θ^* est le mode a posteriori et Σ^{*-1} est définie par :

$$\Sigma^{*-1} = \frac{\partial^2 \log\{l(\theta/n)g(\theta/\beta, \sigma)\}}{\partial \theta^T \partial \theta} \Big|_{\theta = \theta^*}$$

il en résulte:

$$\begin{aligned} t_1^{(p)} &= E\left(\sum_{ij} u_i v_j \lambda_{ij} / n; \beta^{(p)}, \sigma^{(p)}\right) \\ &= \sum_{ij} u_i v_j E(\lambda_{ij} / n; \beta^{(p)}, \sigma^{(p)}) \\ &= \sum_{ij} u_i v_j \lambda_{ij}^* \end{aligned}$$

où λ_{ij}^* est l'estimateur de Bayes pour λ_{ij} obtenu à la $p^{\text{ième}}$ itération.

$$\begin{aligned} t_2^{(p)} &= E\left(\sum_{ij} \lambda_{ij}^2 / n; \beta^{(p)}, \sigma^{(p)}\right) \\ &= \sum_{ij} \left[\text{var}(\lambda_{ij} / n; \beta^{(p)}, \sigma^{(p)}) + E^2(\lambda_{ij} / n; \beta^{(p)}, \sigma^{(p)}) \right] \\ &= \sum_{ij} (\lambda_{ij}^*)^2 + \sum_{k=1}^r \sigma_{kk}^* \end{aligned}$$

où les σ_{kk}^* sont les derniers rc éléments diagonaux de Σ^* .

Ainsi l'étape E de l'algorithme EM est exécutée en prenant:

$$\begin{aligned} t_1^{(p)} &= \sum_{i=1}^r \sum_{j=1}^c u_i v_j \lambda_{ij}^* \\ t_2^{(p)} &= \sum_{i=1}^r \sum_{j=1}^c (\lambda_{ij}^*)^2 + \sum_{k=1}^r \sigma_{kk}^* \end{aligned}$$

Remarque sur l'algorithme EM :

L'algorithme EM appliqué à notre problème converge extrêmement lentement. De plus, la solution n'est pas nécessairement unique. En fait, les couples $(\sigma = 0, \beta \text{ arbitraire})$ sont toujours des solutions.

b) Autre méthode approximative pour estimer (β, σ^2)

On a vu précédemment que la loi de probabilité marginale était reliée à la densité a posteriori par la relation :

$$m(n; \beta, \sigma) = \frac{l(\theta/n)g(\theta/\beta, \sigma)}{h(\theta/n; \beta, \sigma)} \quad \text{pour tout } \theta \quad (3)$$

On peut approcher $m(n; \beta, \sigma)$ en divisant le produit de la vraisemblance et de la densité a priori par l'approximation normale de la densité a posteriori utilisée dans le paragraphe précédent. L'approximation qui en résulte pour la distribution marginale est raisonnable lorsque

l'approximation normale pour la loi a posteriori est évaluée en $\theta = \theta^*$. En utilisant cette substitution pour chaque terme de θ dans (3), on calcule numériquement la distribution marginale approchée pour une grille de valeurs de (β, σ) dans le but de déterminer approximativement la valeur de (β, σ) qui maximise la distribution marginale.

Dans cette 2ème méthode, une très large étendue de valeurs de (β, σ) est premièrement utilisée pour déterminer le comportement général de la distribution marginale et voir la région dans laquelle son maximum apparaît. Puis une grille plus raffinée des valeurs de (β, σ) est utilisée pour mieux localiser ce maximum.

Pour les différents exemples étudiés par les auteurs, cette deuxième approche et celle qui utilise l'algorithme EM donnent des résultats très similaires. Comme la convergence est lente avec l'algorithme EM, il s'avère que la seconde approximation est plus utile particulièrement lorsque $\sigma=0$. D'ailleurs, une caractéristique intéressante de l'approche bayésienne empirique est que si l'ajustement par la méthode du maximum de vraisemblance du modèle $L \times L$ donné par :

$$\log m_{ij} = \mu + \lambda_i^x + \lambda_j^y + \beta u_i v_j$$

est exceptionnellement bon, alors la distribution marginale de n est maximisée en une paire (β, σ) pour laquelle $\sigma=0$. Lorsque ceci a lieu, on présume que la valeur de β est identique à l'emv $\hat{\beta}$ pour le modèle $L \times L$. Ce résultat donne une interprétation intéressante à la solution empirique de Bayes :

Si le modèle $L \times L$ s'ajuste extrêmement bien, on peut estimer les proportions de la table de contingence en utilisant les estimateurs des fréquences espérées $\{\hat{m}_{ij}\}$ par la méthode du maximum de vraisemblance appliquée directement à ce modèle.

IV- EXEMPLES

Maintenant, on va appliquer les différentes méthodes discutées précédemment aux données d'un tableau qui croise l'intensité de peur dans les rêves des enfants avec leurs âges. Pour cela, on considère un échantillon de 223 garçons âgés de 5 à 15 ans, et on choisit comme scores $u' = (6; 8,5; 10,5; 12,5; 14,5)$ (les points milieux pour la variable âge) et $\{v_j = j - 3,5\}$.

La table 1 contient les estimateurs des fréquences espérées $\{\hat{m}_{ij}\}$ pour les différentes approches :

- (1) les estimateurs basés sur les proportions d'échantillonnage $\{\hat{m}_{ij} = n_{ij}\}$;
- (2) les estimateurs du maximum de vraisemblance obtenus à partir de l'ajustement par le modèle $L \times L$;
- (3) les estimateurs obtenus à partir de l'approche empirique de BAYES en utilisant la loi a priori normale $N(\beta_{u,v_j}, \sigma^2)$ pour les $\{\lambda_{ij}\}$ dans le modèle loglinéaire saturé.

TABLE -1-

AGE	ESTIMATEUR	INTENSITE DE PEUR DANS LES REVES			
		<i>Moins terrifiants</i>		<i>très terrifiants</i>	
		1	2	3	4
5 - 7	1	7	4	3	7
	2	4.80	3.39	5.23	7.58
	3	5.34	3.51	4.83	7.31
8 - 9	1	10	15	11	13
	2	16.41	9.09	11.01	12.50
	3	14.61	10.74	10.98	12.66
10 - 11	1	23	9	11	7
	2	21.41	9.76	9.73	9.10
	3	22.09	9.51	9.87	8.54
12 - 13	1	28	9	12	10
	2	30.72	11.53	9.46	7.29
	3	29.67	11.02	10.22	8.09
14 - 15	1	32	5	4	3
	2	26.67	8.24	5.57	3.53
	3	28.29	7.22	5.09	3.41

Le modèle $L \times L$ avec les scores indiqués ajuste assez bien ces données.

On trouve $G^2 = 14,6$ pour $dL=11$ et l'emv de β est $\hat{\beta} = -0,097$.

$G^2 = 14,6$ étant le rapport de vraisemblance défini par :

$$G^2 = 2 \sum \sum n_{ij} \log (n_{ij} / \hat{m}_{ij})$$

La valeur négative de $\hat{\beta}$ indique qu'il y a une association négative entre nos deux variables catégorielles. En d'autres termes, l'intensité de peur dans les rêves tend à croître dans le sens inverse de l'âge des garçons. Par conséquent, ce sont les enfants les plus jeunes qui ont tendance à avoir les rêves les plus terrifiants.

Pour l'approche bayésienne empirique, avec la loi a priori normale pour les $\{\lambda_{ij}\}$, l'algorithme EM produit $\beta = -0,092$ et $\sigma = +0,184$.

L'approximation directe de la distribution marginale de n suggère que cette dernière est maximisée pour $\beta = -0,094$ et $\sigma = +0,187$.

La valeur de σ relativement petite comparée à $\beta u_i v_j$ reflète le bon ajustement obtenu avec le modèle $L \times L$.

Les valeurs $\beta = -0,092$ et $\sigma = +0,184$ sont utilisées dans la loi a priori normale qui génère la distribution a posteriori correspondant aux estimateurs des fréquences espérées données dans la table.

Maintenant, on va considérer une autre table qui croise le niveau social et économique des parents avec l'état de santé mentale. Les différents estimateurs des fréquences espérées sont illustrés dans la table 2. Ils sont basés sur :

- (1) Les proportions d'échantillonnage.
- (2) Les estimateurs du maximum de vraisemblance pour le modèle U .
- (3) Les estimateurs empiriques de BAYES.

En fait, cette table est un exemple de données pour lesquelles l'approche bayésienne empirique donne une loi a priori dégénérée ($\sigma = 0$) et donc les estimateurs des $\{\pi_{ij}\}$ sont identiques aux estimateurs du maximum de vraisemblance pour le modèle U .

TABLE -2-

Etat de santé mentale	Estimateur	Niveau social et économique des parents					
		A	B	C	D	E	F
Bon	1	64	57	72	57	36	21
	2	65.3	54.2	55.9	65.3	39.0	27.3
	3	65.3	54.2	55.9	65.3	39.0	27.3
Symptômes légers	1	94	94	105	141	97	71
	2	104.4	94.4	107.2	137.0	89.6	68.8
	3	104.4	94.4	107.2	137.0	89.6	68.8
Symptômes modérés	1	58	54	65	77	54	54
	2	50.2	49.9	61.7	86.4	61.8	52.0
	3	50.2	49.9	61.7	86.4	61.8	52.0
Grave	1	46	40	60	94	78	71
	2	42.1	45.9	62.2	95.3	74.7	68.8
	3	42.1	45.9	62.2	95.3	74.7	68.8

Le modèle U ajuste les données extrêmement bien. D'ailleurs, on trouve $G^2 = 9,9$ pour $dL = 14$ et l'emv de β est: $\hat{\beta} = 0,091$.

Ici, l'association entre les deux variables est positive et uniforme. Par conséquent, l'étude de cet exemple suggère que l'état de santé mentale s'améliore avec le niveau social et économique des individus.

L'algorithme EM et l'approximation directe de la marginale de n donnent des résultats similaires :

$\hat{\beta} = 0,09$ et $\sigma = 0$ pour la distribution a priori.

REFERENCES

A. AGRESTI (1990)

Categorical Data Analysis, New-York; John Wiley.

A.P. DEMPSTER, N.M. LAIRD & D.B. RUBIN (1977)

Maximum likelihood from incomplete data via the EM algorithm (with discussion), JRSS B 39, pp 1-38

L. GOODMAN (1979)

Simple models for the analysis of association in cross classifications having ordered categories, JASA 74, pp. 537-552

N.M. LAIRD (1978)

Empirical Bayes methods for two-way contingency tables, Biometrika 65pp. 581-590.