C. BASTIN

Discrimination par ajustement à des lois normales multidimensionnelles : exemple d'une distribution en anneau

Les cahiers de l'analyse des données, tome 2, nº 4 (1977), p. 407-411

<http://www.numdam.org/item?id=CAD_1977__2_4_407_0>

Les Cahiers de l'Analyse de Données
Vol. II - 1977 - n° 4 - p. 407-411

DISCRIMINATION PAR AJUSTEMENT A DES LOIS NORMALES MULTIDIMENSIONNELLES : EXEMPLE D'UNE DISTRIBUTION EN ANNEAU [DISCR. ANNEAU]

par C. Bastin (1)

1 Enoncé du problème

Il est d'usage pour nombre de statisticiens, d'assimiler toute distribution empirique de points, ou toute loi de probabilité, à la loi normale ayant même centre et même variances et covariances. A partir de celà le problème de la discrimination - placer une cloison entre deux distributions de masses qui représentent e.g. deux catégories de malades se résoudra en traçant la ligne où sont égales les densités de deux lois normales ; etc...(Sur cette méthode cf [Sep. Corr.] §4 dans ce cahier p 383)

L'objet de ce problème est d'examiner le succès de ce procédé, sur un exemple parfois cité en sa faveur.(cf[Sep. Corr.] § 4' Note ; p 385)

Dans tout le problème on se bornera au cas de distributions planes, et on notera :

$$N_\sigma(x,y)\ dx\ dy = (2\pi)^{-1}\ \sigma^{-2}\ \exp(-(x^2+y^2)/(2\sigma^2))\ dx\ dy$$
$$= (2\pi)^{-1}\ \sigma^{-2}\ \exp(-\rho^2/(2\sigma^2))\ \rho\ d\rho\ d\theta$$

(où $\rho^2 = x^2+y^2$; ρ , θ sont les coordonnées polaires usuelles).

1.1 Calculer directement pour la loi $N_\sigma(x,y)\ dx\ dy$, l'espérance mathématique

$$Esp_{N\sigma}(\rho^2) = Esp_{N\sigma}(x^2+y^2)$$

En déduire les valeurs de $Esp_{N\sigma}(x^2)$ et $Esp_{N\sigma}(y^2)$.

1.2 On considère la loi définie par la formule :

$$D_a(x,y) = \text{si } x^2 + y^2 < a^2 \text{ alors } (1/(\pi a^2)) \text{ sinon } 0$$

i.e. densité constante dans le disque de rayon a ; et masse totale 1

Calculer pour cette loi $D_a(x,y)\ dx\ dy$, l'espérance mathématique

$$Esp_{Da}(\rho^2).$$

En déduire la valeur de $\sigma(a)$ telle que $N_{\sigma(a)}$ soit la loi ajustée à D_a, c'est-à-dire la loi normale ayant même centre et mêmes variances et covariances que D_a.

1.3 On considère la loi définie par la formule :

$$C_{b,c}(x,y)=\text{si } b^2 < x^2 + y^2 < c^2 \text{ alors } (1/(\pi(c^2 - b^2))) \text{ sinon } 0$$

(i.e. densité constante sur la couronne de rayon intérieur b, rayon extérieur c ; et masse totale 1).

Calculer pour cette loi $C_{b,c}(x,y)\ dx\ dy$, l'espérance mathématique

$$EspC_{b,c}(\rho^2)$$

---

(1) Assistante de statistique à l'Université Pierre et Marie Curie, Paris

En déduire la valeur de $\sigma(b,c)$ telle que $N_{\sigma(b,c)}$ soit la loi ajustée à $C_{b,c}$, c'est-à-dire la loi normale ayant même centre et mêmes variances covariances que $C_{b,c}$.

Ces deux lois vous paraissent-elles se ressembler ?

1.4 On suppose désormais que $a < b < c$. Montrer qu'il existe $\rho(a,b,c)$ tel que l'on ait :

1) si $x^2 + y^2 < \rho^2(a,b,c)$: $N_{\sigma(a)}(x,y) > N_{\sigma(b,c)}(x,y)$

2) si $x^2 + y^2 = \rho^2(a,b,c)$: $N_{\sigma(a)}(x,y) = N_{\sigma(b,c)}(x,y)$

3) si $x^2 + y^2 > \rho^2(a,b,c)$: $N_{\sigma(a)}(x,y) < N_{\sigma(b,c)}(x,y)$

Exprimer en fonction de $t = a^2/(b^2 + c^2)$ la valeur de $(\rho^2(a,b,c)/a^2)$.

On notera $\rho^2(a,b,c)/a^2 = \varphi(t)$.

1.5 Selon le principe de discrimination posé en introduction, le cercle de rayon $\rho(a,b,c)$ doit se placer entre le disque D_a et la couronne $\mathcal{C}_{b,c}$. Mais cela n'est pas vrai pour toutes valeurs de a,b,c. Dans les questions 5° à 8° on se propose de chercher à quelle condition on a vraiment :

$$a < \rho(a,b,c) < b$$

En se souvenant de la condition $a < b < c$, tracer la partie utile de la courbe représentative de la fonction $\varphi(t)$ et en tirer une condition pour que $a < \rho(a,b,c)$.

1.6 Soit $t < 0{,}5$. Quelle est la plus petite valeur de (b/a) compatible avec les conditions :

$$t = a^2/(b^2 + c^2) \ ; \ a < b < c.$$

En supposant que (b/a) ait cette valeur minimum, est-il possible que l'on ait :

$$a \leq \rho(a,b,c) \leq b$$

Se servant de la question précédente préciser quelles sont alors les valeurs de (b/a) et de (c/a).

1.7 Soit $t < 0{,}5$. Quelle est la plus grande valeur de (b/a) compatible avec les conditions :

$$t = a^2/(b^2 + c^2) \ ; \ a < b < c.$$

Quelle est alors la forme de la couronne $C_{b,c}$?

En supposant que (b/a) ait cette valeur maxima, exprimer en fonction de t le rapport $\rho^2(a,b,c)/b^2 = \psi(t)$.

1.8 En traçant la courbe représentative de la fonction $\psi(t)$ préciser à quelle condition en t, on a $\rho(a,b,c) < b$, sous l'hypothèse de 7° (i. e. (b/a) maximum pour t fixé).

N.B.: La première moitié de la question 6°, ainsi que la question 7° tout entière se peuvent traiter avant la question 5°.

En 6° et 7° "soit $t < 0{,}5$" est à comprendre "soit t fixé et $t < 0{,}5$"

2 *Solution du problème*

2.1 $Esp_{N\sigma}(\rho^2) = 2\ \sigma^2$

donc $Esp_{N\sigma}(x^2) = Esp_{N\sigma}(y^2) = (1/2)\ Esp_{N\sigma}(x^2 + y^2) = \sigma^2$

2.2 $Esp_{Da}(\rho^2) = a^2/2$

donc $Esp_{Da}(x^2) = Esp_{Da}(y^2) = (1/2) Esp_{Da}(x^2 + y^2) = a^2/4$.

Les lois D_a et N_σ sont centrées ; et pour ces deux lois, toutes les covariances sont nulles.

Le nombre $\sigma(a)$ tel que $N_{\sigma(a)}$ soit la loi normale ajustée à D_a est donc $\sigma(a) = a/2$

2.3 $Esp\ \mathcal{C}_{b,c}(\rho^2) = (b^2 + c^2)/2$

$Esp\ \mathcal{C}_{b,c}(x^2) = Esp\ \mathcal{C}_{b,c}(y^2) = (b^2 + c^2)/4$

La loi $\mathcal{C}_{b,c}$ est centrée et ses covariances sont nulles.

Le nombre $\sigma(b,c)$ tel que $N_{\sigma(b,c)}$ soit la loi normale ajustée à $\mathcal{C}_{b,c}$ est $\sigma(b,c) = (b^2 + c^2)^{1/2}/2$

Remarque : Les deux lois $N_{\sigma(b,c)}$ et $\mathcal{C}_{b,c}$ sont pourtant très différentes : la première est concentrée autour de son centre, alors que la seconde est nulle autour de son centre. C'est ce qu'on voit sur la figure illustrant le § 4' (Note p 385) de [Sep. Corr.]

2.4 $N_{\sigma(a)}\ (x,y) \geq N_{\sigma(b,c)}\ (x,y)$

équivaut à

$$\sigma(a)^{-2} \exp(-\ \rho^2/(2\sigma(a)^2)) \geq \sigma(b,c)^{-2} \exp(-\ \rho^2/(2\sigma\ (b,c)^2))$$

$$\Leftrightarrow \quad \exp(-\ 2\ \rho^2/\ a^2) \geq (a^2/(b^2 + c^2)) \exp(-\ 2\ \rho^2/(b^2 + c^2))$$

$$\Leftrightarrow \quad -\ 2\ \rho^2/\ a^2 \geq \text{Logt} - 2\ \rho^2/(b^2 + c^2)$$

$$\Leftrightarrow \quad \rho^2/\ a^2 \leq -\ \text{Logt}/(2(1 - t))$$

$$\text{où} \quad t = a^2/(b^2 + c^2)$$

donc si $\rho(a,b,c)$ est le nombre tel que

$$\rho^2(a,b,c)/a^2 = \text{Logt}/(2(1 - t)) = \varphi(t),$$

alors les conditions 1,2,3 sont satisfaites.

C'est ce nombre $\rho(a,b,c)$ qui permet, d'après le principe posé en introduction, de discriminer les lois $N_{\sigma(a)}$ et $N_{\sigma(b,c)}$. Le but des questions suivantes est de regarder, pour certaines valeurs de a,b,c, si cette solution est satisfaisante pour la discrimination des lois D_a et $\mathcal{C}\ b,c$.

2.5 $\varphi(t) = -\text{Log}t/(2(1-t))$

$\varphi'(t) = -(t\,\text{Log}t + 1 - t)/(2t(1-t)^2)$

puisque $a < b < c$, $0 < t < 1/2$.

L'étude de $u(t) = t\,\text{Log}t + 1 - t$ permet de montrer que $\varphi'(t)$ est toujours négative et donc que $\varphi(t)$ est décroissante (cf figure 1).

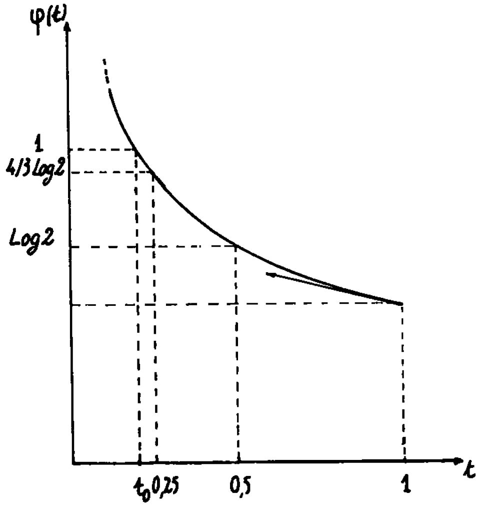

Figure 1 : courbe représentative de la fonction $\varphi(t) = -\text{Log}\,t/(2(1-t))$.
$\varphi(0,2) > 1 > \varphi(0,205)$
$0,2 < t_0 < 0,205$

La condition $a \le \rho(a,b,c)$, qui est équivalente à la condition $\varphi(t) \ge 1$, est réalisée dès que $0 < t \le t_0$, si t_0 désigne l'unique valeur de t telle que $\varphi(t_0) = 1$. ($t_0 \triangleq 0,2$)

Remarque : Pour toutes les valeurs de a,b,c telles que $t > t_0$, i.e $\rho(a,b,c) < a$, la discrimination entre les lois D_a et $C_{b,c}$, à l'aide du cercle de rayon $\rho(a,b,c)$ et de centre O, utilisé comme cloison de séparation n'est pas satisfaisante ; en effet cette cloison ne sépare pas parfaitement les domaines de définition de D_a et $C_{b,c}$ qui sont disjoints, alors que tout cercle de centre O et de rayon compris entre a et b assure une séparation parfaite.

2.6 Comme $t = a^2/(b^2 + c^2) < 0,5$ et $a < b < c$, la valeur minimale de b/a compatible avec ces conditions est $b/a = 1$.

Donc la seule valeur de $\rho(a,b,c)$ vérifiant

$$a \le \rho(a,b,c) \le b \text{ est } \rho(a,b,c) = a = b.$$

et alors $\varphi(t) = 1$ et $t = t_0 \triangleq 0,2$.

$$b/a = 1 \text{ et } c/a = (1/t_0 - 1)^{1/2} \triangleq 2.$$

Donc si $a = b$, alors que le cercle de centre O et de rayon a sépare parfaitement D_a et $C_{b,c}$, le cercle de centre O et de rayon $\rho(a,b,c)$ ne sépare pas parfaitement D_a et $C_{b,c}$, sauf pour la valeur $c=a(1/t_0-1)^{1/2}$, auquel cas $\rho(a,b,c) = a$. La discrimination obtenue à l'aide du cercle de centre O et de rayon $\rho(a,b,c)$ n'est donc pas satisfaisante.

2.7 La plus grande valeur de b/a compatible avec les conditions $t = a^2/(b^2 + c^2) < 0,5$ et $a < b < c$ est $b/a = (1/2\,t)^{1/2} = c/a$.

La couronne $C_{b,c}$ est réduite à un cercle.

(En réalité la loi $C_{b,c}$ n'est pas définie dans ce cas-là ; mais on peut considérer des cas proches de ce cas limite : $c = b + \varepsilon$, où ε est une quantité positive infiniment petite)

$$\psi(t) = \rho^2(a,b,c)/b^2 = (a^2/b^2)\ \varphi(t)$$

$$\psi(t) = -(t \operatorname{Log} t)\ /(1 - t)$$

2.8 $$0 < t < 0,5$$

$$\psi(t) = -(t \operatorname{Log} t)\ /(1 - t)$$

$$\psi'(t) = -(\operatorname{Log} t + 1 - t)/(1 - t)^2$$

L'étude de $\psi'(t)$ montre que $\psi(t)$ est croissante et toujours inférieure à 1.

Donc $\rho(a,b,c)$ est toujours inférieur à b, mais n'est supérieur à a que pour $t < t_0$. (cf figure 2).

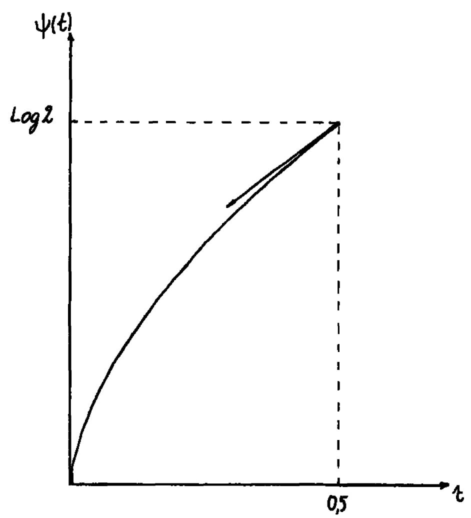

Figure 2: courbe représentative de la fonction $\psi(t) = -t \operatorname{Log} t\ /(1-t)$.