

ANNIE MORIN

## A propos de la loi de Poisson multivariée

*Journal de la société statistique de Paris*, tome 134, n° 2 (1993), p. 3-13

[http://www.numdam.org/item?id=JSFS\\_1993\\_\\_134\\_2\\_3\\_0](http://www.numdam.org/item?id=JSFS_1993__134_2_3_0)

© Société de statistique de Paris, 1993, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## COMMUNICATIONS

**A propos de la loi de Poisson multivariée**

Annie MORIN \*  
ENSAE / IRISA

**Abstract**

Multidimensional data which are skewed, kurtotic, multimodal or zero-one on one or more dimensions lead to questions about accuracy and reliability of classical gaussian discriminant analysis with non-gaussian data. There are two types of approach to changing the gaussian hypothesis, one to adopt some other parametric hypothesis which seems more appropriate to the data at hand, and the other is to pursue a more basic non-parametric attitude. Actually, skewness being often the most striking non-gaussian aspect of the data, and the Poisson being the archetypical skewed discrete distribution, the parametric chosen here for study is the Poisson hypothesis. We review in some detail the bivariate and multivariate Poisson distribution and we give some strategies employed to use this knowledge in classifying data by the maximum likelihood method.

**1. Introduction**

Les techniques statistiques utilisées en reconnaissance des formes permettent de prendre une décision de reconnaissance sur un objet ou un individu. La plupart du temps, cette décision se traduit par une affectation de l'objet sur lequel on a effectué des mesures à un groupe ou une classe préspecifiée. Ces méthodes statistiques s'appuient en général sur des hypothèses concernant la description de ces classes d'objets. On suppose que les mesures faites sur l'objet à reconnaître peuvent s'écrire sous la forme d'un vecteur  $X$  de  $R^p$ . On dispose d'un ensemble d'apprentissage, c'est-à-dire d'un ensemble d'observations sur des objets pour lesquels on connaît la classe d'appartenance. Cet ensemble d'apprentissage va permettre de déterminer des règles d'affectation pour de nouvelles observations dont on ignore la classe. Donc, en

---

\* Exposé fait le 20 janvier 1992, à l'occasion de la remise du Prix 1991 du Statisticien d'expression française décerné par la S.S.P.

## A PROPOS DE LA LOI DE POISSON MULTIVARIÉE

résumé, un système de reconnaissance des formes peut être considéré comme une règle de décision automatique en ce sens qu'il transforme un vecteur de mesure en une affectation à une classe.

Les formes à reconnaître sont elles-mêmes très variées. Il peut s'agir de reconnaissance de pièces, d'analyse d'images de satellites ou de phonèmes en reconnaissance de la parole.

La difficulté de la reconnaissance est due, d'une part au bruit qui perturbe les mesures sur les objets, et d'autre part, à la grande variabilité des objets appartenant à une même classe. Par ailleurs, les classes définies *a priori* sont souvent des classes naturelles, par exemple les sons élémentaires en parole. Elles peuvent se révéler mal adaptées à l'espace de représentation qu'on a choisi. En particulier, elles peuvent se recouvrir ou mal décrire l'ensemble d'apprentissage.

L'approche statistique n'est pas la seule méthodologie utilisée en reconnaissance des formes. Elle a l'avantage de se placer dans un cadre mathématique solide mais l'inconvénient d'oublier la nature des mesures faites sur les objets. L'approche structurelle (Miclet, 1984) s'attache à définir les caractéristiques intrinsèques de la forme plutôt qu'à donner sa description métrique, c'est-à-dire le vecteur de mesures (coordonnées) des objets. En pratique, ces deux approches sont complémentaires et on a tout intérêt à les utiliser toutes les deux.

Ici, nous nous intéressons uniquement à l'approche statistique et plus précisément au modèle de Poisson multidimensionnel. La loi de Poisson multidimensionnelle est beaucoup moins connue que la loi normale multidimensionnelle. C'est pourquoi nous la présentons de façon détaillée. Nous comparons ensuite les performances des modèles poissonniens et gaussiens sur des données simulées et sur des images de satellite Landsat.

## 2. La loi de Poisson bivariée

**Définition :** Une distribution discrète multivariée est de Poisson si ses marginales suivent des lois de Poisson unidimensionnelles.

### 2.1 Structure d'une loi de Poisson bivariée

La distribution de Poisson bivariée a été introduite en 1934 par Campbell qui a considéré la limite de la distribution d'une table de contingence à deux dimensions. La forme explicite de la loi est due quelques années plus tard à Aitken [1944]. Pratiquement, à la même époque, Guldberg [1934] obtient la distribution bivariée de lois de Poisson indépendantes comme limite de la distribution de lois binomiales indépendantes. Il faut cependant attendre Holgate en 1964 pour obtenir une variable de Poisson bivariée à partir de trois variables univariées de Poisson indépendantes, c'est-à-dire avec une matrice de variance-covariance non diagonale.

## A PROPOS DE LA LOI DE POISSON MULTIVARIÉE

Notons  $X_1, X_2$  et  $X_3$  trois variables de Poisson indépendantes de paramètres  $\lambda_1, \lambda_2$  et  $\lambda_3$  respectivement et considérons la distribution conjointe de  $X = X_1 + X_2$  et de  $Y = X_1 + X_3$ .

$$P[X = k, Y = l] = e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{i=0}^{\min(k, l)} \frac{\lambda_1^i}{i!} \frac{\lambda_2^{k-i}}{(k-i)!} \frac{\lambda_3^{l-i}}{(l-i)!} \quad (1)$$

$$k = 0, 1, \dots$$

$$l = 0, 1, \dots$$

La covariance de  $X$  et  $Y$  est  $\text{cov}(X, Y) = \lambda_1$  et la corrélation entre  $X$  et  $Y$  s'écrit

$$\rho_{XY} = \frac{\lambda_1}{(\lambda_1 + \lambda_2)^{1/2} (\lambda_1 + \lambda_3)^{1/2}} \quad (2)$$

Plus récemment, Kawamura [1973] a considéré la structure d'une loi de Poisson bivariée comme limite d'une distribution bivariée de Bernoulli et a retrouvé les résultats de Holgate.

Soient  $X$  et  $Y$  deux variables de Bernoulli dont la distribution conjointe est donnée par :

$$P(X = i, Y = j) = P_{ij} \quad i = 0, 1 \quad j = 0, 1 \quad (3)$$

Les probabilités marginales pour  $X$  et  $Y$  sont

$$\begin{aligned} P[X = 0] &= P_{00} + P_{01} & P[Y = 0] &= P_{00} + P_{10} \\ P[X = 1] &= P_{10} + P_{11} & P[Y = 1] &= P_{01} + P_{11} \end{aligned}$$

La fonction génératrice d'une distribution de Bernoulli est

$$\begin{aligned} g(s_1, s_2) &= \sum_{i=0,1; j=0,1} P_{ij} s_1^i s_2^j \\ &= P_{00} + P_{10} s_1 + P_{01} s_2 + P_{11} s_1 s_2 \end{aligned} \quad (4)$$

Supposons que les événements  $(0, 0), (1, 0), (0, 1)$  and  $(1, 1)$  se produisent  $\alpha, \beta, \gamma$  and  $\delta$  fois, respectivement avec  $\alpha + \beta + \gamma + \delta = n$ . Alors

$$\begin{aligned} P \left[ \sum_{i=1}^r X_i = \beta + \delta, \sum_{i=1}^r Y_i = \gamma + \delta \right] &= \frac{n!}{\alpha! \beta! \gamma! \delta!} P_{00}^\alpha P_{10}^\beta P_{01}^\gamma P_{11}^\delta \\ &= P_{\alpha\beta\gamma\delta} \end{aligned} \quad (5)$$

La fonction génératrice de  $\left( \sum_{i=1}^n X_i, \sum_{i=1}^n Y_i \right)$  est donnée par  $g(s_1, s_2)^n$ .

Notons au passage que les distributions marginales de

$$\sum_{i=1}^n X_i \quad \text{et} \quad \sum_{i=1}^n Y_i$$

## A PROPOS DE LA LOI DE POISSON MULTIVARIÉE

sont

$$P \left[ \sum_{i=1}^n X_i = k \right] = \sum_{\beta + \delta = k; \alpha + \beta + \gamma + \delta = n} P_{\alpha\beta\gamma\delta} = C_n^k (P_{10} + P_{11})^k (P_{00} + P_{01})^{n-k} \quad (6)$$

$0 \leq k \leq n$

i.e.

$$\sum_{i=1}^n X_i \simeq \mathbf{B}(n, P_{10} + P_{11})$$

$$P \left[ \sum_{i=1}^n Y_i = l \right] = C_n^l (P_{01} + P_{11})^l (P_{00} + P_{10})^{n-l} \quad (7)$$

$0 \leq l \leq n$

i.e.

$$\sum_{i=1}^n Y_i \simeq \mathbf{B}(n, P_{10} + P_{11})$$

La covariance entre  $X$  et  $Y$  s'écrit

$$\text{cov}(X, Y) = P_{00}P_{11} - P_{10}P_{01} \quad (8)$$

tandis que la covariance entre  $\sum_{i=1}^n X_i$  et  $\sum_{i=1}^n Y_i$  est :

$$\text{cov} \left( \sum_{i=1}^n X_i, \sum_{i=1}^n Y_i \right) = n (P_{00}P_{11} - P_{10}P_{01}) \quad (9)$$

Kawamura prend la limite de la binomiale bivariable lorsque  $n \rightarrow \infty$  et  $P_{ij} = \frac{\lambda_j}{n}$  où  $\lambda_j$  est fini et fixé, par analogie avec la situation univariée et retrouve les résultats de Holgate.

### 2.2 Les propriétés de la loi de Poisson bivariable

On a :

$$P[X = k] = e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!}$$

$$P[Y = l] = e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^l}{l!}$$

$$p_{kl} = P[X = k, Y = l] = e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{i=0}^{\min(k, l)} \frac{\lambda_1^i}{i!} \frac{\lambda_2^{k-i}}{(k-i)!} \frac{\lambda_3^{l-i}}{(l-i)!}$$

## A PROPOS DE LA LOI DE POISSON MULTIVARIÉE

### Relations de récurrence

Pour calculer les valeurs numériques de  $p_{kl}$ , on a besoin des relations de récurrence suivantes [Tei54] :

$$\begin{aligned}kp_{kl} &= \lambda_2 p_{k-1, l} + \lambda_1 p_{k-1, l-1} \\lp_{kl} &= \lambda_3 p_{k, l-1} + \lambda_1 p_{k-1, l-1}, \quad k \geq 0, l \geq 0\end{aligned}$$

Si  $k$  ou  $l$  est négatif,  $P_{kl} = 0$ .

### Fonction caractéristique

Dwass et Teicher ont montré qu'une distribution infiniment divisible avec des marginales de Poisson doit avoir une fonction caractéristique de la forme :

$$\exp \left( \sum_{j=1}^n a_j E_j + \sum_{i < j} a_{ij} E_i E_j + \dots + a_{12\dots n} \sum_{j=1}^n E_j - A_n \right) \quad (10)$$

où  $E_j = \exp(i, S_j)$  et  $A_n$  est une constante telle que (10) = 1 si  $E_j = 1$ .

i.e.

$$A_n = \sum_{j=1}^n a_j + \sum_{i < j} a_{ij} + \dots + a_{12\dots n}$$

### Lois conditionnelles

$$\begin{aligned}P(X = k/Y = 1) &= P_{kl} \left[ e^{-(\lambda_1 + \lambda_3)} \frac{(\lambda_1 + \lambda_3)^l}{l!} \right]^{-1} \\&= e^{-\lambda_2} \sum_{j=0}^{\min(k, l)} C_l^j \left( \frac{\lambda_1}{\lambda_1 + \lambda_3} \right)^j \left( \frac{\lambda_3}{\lambda_1 + \lambda_3} \right)^{l-j} \frac{\lambda_2^{k-j}}{(k-j)!}\end{aligned} \quad (11)$$

$$E(X/Y) = \lambda_2 + l \frac{\lambda_1}{\lambda_1 + \lambda_3} \quad (12)$$

$$\text{var}(X/Y = 1) = \lambda_2 + l\lambda_1\lambda_3/(\lambda_1 + \lambda_3)^2 \quad (13)$$

La loi conditionnelle de  $X$  sachant  $Y$  est la convolution de  $U$  poissonnienne de paramètre  $\lambda_2$  et de  $V$  binomiale de paramètres  $l$  et  $\frac{\lambda_1}{\lambda_1 + \lambda_3}$ .

$$\begin{aligned}\rho_{xy} &= \frac{\lambda_1}{\sqrt{\lambda_1 + \lambda_3} \sqrt{\lambda_1 + \lambda_2}} \leq \lambda_1 (\lambda_1 + \min(\lambda_2, \lambda_3))^{-1} \\&\leq \left( \frac{\lambda_1 + \lambda_3}{\lambda_1 + \lambda_2} \right)^{1/2}, \quad \left( \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_3} \right)^{1/2}\end{aligned} \quad (14)$$

## A PROPOS DE LA LOI DE POISSON MULTIVARIÉE

Enfin, le développement en série est moins utilisé pour les distributions discrètes que continues. Cependant, pour une Poisson univariée, on a les développements en polynômes de Charlier de type B. Ceci se généralise naturellement au cas bivarié.

$$P [X = k, Y = l] = \left( e^{-\theta_1} \frac{\theta_1^k}{k!} \right) \left( e^{-\theta_2} \frac{\theta_2^l}{l!} \right) \left( 1 + \frac{\bar{\theta}}{1!} G_1(k, \theta_1) \times G_1(l, \theta_2) + \frac{\bar{\theta}^2}{2!} G_2(k, \theta_1) G_2(l, \theta_2) + \dots \right) \quad (15)$$

où

$$\theta_1 = E(X)$$

$$\theta_2 = E(Y)$$

$$\bar{\theta} = \rho_{xy} (\theta_1 \theta_2)^{1/2}$$

et les  $G_1$  sont les polynômes de Charlier de type B.

### 2.3 Estimation des paramètres

Rappelons que

$$P [X = k, Y = l] = e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{i=0}^{\min(k, l)} \frac{\lambda_1^i}{i!} \frac{\lambda_2^{k-i}}{(k-i)!} \frac{\lambda_3^{l-i}}{(l-i)!} \quad (16)$$

$$k = 0, 1, \dots$$

$$l = 0, 1, \dots$$

avec  $E(X) = \lambda_1 + \lambda_2 = \theta_1$

et  $E(Y) = \lambda_1 + \lambda_3 = \theta_2$

$$\text{var}(X) = \lambda_1 + \lambda_2 = \theta_1$$

et  $\text{var}(Y) = \lambda_1 + \lambda_3 = \theta_2$

$$\text{cov}(X, Y) = \lambda_1.$$

La méthode des moments permet d'estimer  $\theta_1$ ,  $\theta_2$  et  $\lambda_1$  à partir des moyennes empiriques marginales et de la covariance empirique.

On peut donc réécrire l'équation précédente et utiliser la méthode du maximum de vraisemblance.

$$P [X = k, Y = l] = e^{-(\theta_1 + \theta_2 - \lambda_1)} \sum_{i=0}^{\min(k, l)} \frac{\lambda_1^i}{i!} \frac{(\theta_1 - \lambda_1)^{k-i}}{(k-i)!} \frac{(\theta_2 - \lambda_1)^{l-i}}{(l-i)!} \quad (17)$$

Pour estimer  $\theta_1$ ,  $\theta_2$ ,  $\lambda_1$ , on dérive et on obtient le système d'équations suivantes :

A PROPOS DE LA LOI DE POISSON MULTIVARIÉE

$$\begin{aligned}\frac{\partial P_{kl}}{\partial \theta_1} &= P_{k-1,l} - P_{k,l} \\ \frac{\partial P_{k,l}}{\partial \theta_2} &= P_{k,l-1}\end{aligned}\quad (18)$$

$$\frac{\partial P_{k,l}}{\partial \lambda_1} = P_{k,l} - P_{k,l-1} - P_{k-1,l} + P_{k-1,l-1}$$

$$\sum \frac{1}{P_{k,l}} \frac{\partial P_{k,l}}{\partial \theta_1} = \sum \frac{1}{P} \frac{\partial P}{\partial \theta_2} = \sum \frac{1}{P} \frac{\partial P}{\partial \lambda_1} = 0 \quad (19)$$

devient

$$\begin{cases} \frac{\bar{x}}{\theta_1 - \lambda_1} - \frac{\lambda_1}{\theta_1 - \lambda_1} \bar{R} - 1 = 0 \\ \frac{\bar{y}}{\theta_2 - \lambda_1} - \frac{\lambda_1}{\theta_2 - \lambda_1} \bar{R} - 1 = 0 \\ \frac{\bar{x}}{\theta_1 - \lambda_1} + \frac{\bar{y}}{\theta_2 - \lambda_1} - \left(1 + \frac{\lambda_1}{\theta_1 - \lambda_1} + \frac{\lambda_1}{\theta_2 - \lambda_1}\right) \bar{R} - 1 = 0 \end{cases} \quad (20)$$

avec

$$R(k, 1) = \frac{P_{k-1,l-1}}{P_{k,l}} \bar{R} = \frac{1}{n} \sum_{k,l} R(k, l) \text{ observé.}$$

On peut donc utiliser les estimations :

$$\begin{aligned}\hat{\theta}_1 &= \bar{x} \\ \hat{\theta}_2 &= \bar{y}\end{aligned}$$

Si  $\theta_1 = \theta_2$ , puis  $\hat{\theta} = \frac{1}{2}(\bar{x} + \bar{y})$ .

L'équation  $\bar{R} = 1$  peut être résolue en utilisant des méthodes itératives. On a aussi :

$$\text{var } \hat{\lambda}_1 = \frac{\lambda_1(Q-1) - 1(\lambda_2 + \lambda_3)(\lambda_2\lambda_3 + \lambda_1)}{[(Q-1-\lambda_2+\lambda_3)(\lambda_1+\lambda_2)(\lambda_1+\lambda_3)]}$$

$$Q = \sum_{k,l=1}^{\infty} \frac{P_{k,l-1}^2}{P_{k,l}}$$

Alternativement,  $\lambda_1$  peut être estimé par la covariance empirique. Mais dans ce cas, la variance de l'estimateur est :

$$\text{var } \hat{\lambda}_1 \approx \left[ \lambda_1(1 + \lambda_1) + (\lambda_1 + \lambda_2)(\lambda_1 + \lambda_3) \right] / n.$$

Cet estimateur n'est corrélé ni à  $\hat{\theta}_1$ , ni  $\hat{\theta}_2$ .



### 3. La loi de Poisson multivariée

Contrairement aux apparences et à ce que disent certains auteurs, la généralisation au cas multivarié n'est pas évidente. Les fonctions génératrices sont de la forme :

$$\exp \left( \sum_{j=1}^K A_j t_j + \sum_{i < j} \Sigma A_{ij} t_i t_j + \dots + A_{12\dots K} t_1 \dots t_K - A \right) \quad (21)$$

où

$$\sum_{j=1}^K A_j + \sum_{i < j} \Sigma A_{ij} + \dots + A_{12\dots K} = A$$

La forme explicite sera :

$$P [X_1 = K_1, X_2 = K_2, \dots, X_m = K_m] = \prod_{j=1}^m e^{-\theta_j} \frac{\theta_j^{K_j}}{K_j!} \quad (22)$$

$$\exp \left( \sum_i \sum_j \theta_{ij} G(K_i) G(K_j) + \sum_i \sum_j \sum_k \theta_{ijk} G(K_i) G(K_j) G(K_k) + \dots + \theta_{12\dots m} \prod_{i=1}^m G(K_i) \right)$$

où les  $G(l)$  sont des polynômes de Charlier de type B.

$$\theta_{ijkl\dots} = E(X_i X_j \cdot X_k \dots)$$

### 4. L'implémentation

La loi de Poisson bivariée peut être simulée à partir de trois variables de Poisson univariées indépendantes. En reconnaissance des formes tout comme en analyse discriminante décisionnelle, un nouvel individu est affecté à une des  $k$  classes  $d$ .

L'expression de la loi de Poisson multivarié est compliquée, il faut donc utiliser les développements en polynômes, soit des approximations, soit les deux simultanément.

Une façon de procéder consiste à utiliser l'approximation de Bahadur [1961] développée pour des variables de Bernoulli :

Soit  $X = (X_1, \dots, X_p)$ ,

$$\theta_j = P(X_j = 1), \quad 1 - \theta_j = P(X_j = 0), \quad z_j = \frac{X_j - \theta_j}{\sqrt{\theta_j(1 - \theta_j)}} \quad (23)$$

où  $X_j$  est 0 ou 1. Bahadur décompose la probabilité conjointe en un produit de deux termes : le premier étant les probabilités dans le cas d'indépendance et le second égal à  $p(x)$ .

## A PROPOS DE LA LOI DE POISSON MULTIVARIÉE

$$p(x) = 1 + \sum_{j < k} \rho_{jk} z_j z_k + \dots + \rho_{12\dots p} z_1 \dots z_p \quad (24)$$

avec

$$\rho_{jk} = E(z_j z_k)$$

$$\rho_{12\dots p} = E(z_1 \dots z_p)$$

corrélation entre  $x_j$  et  $x_k$ .

On peut espérer que dans beaucoup d'applications les interactions d'ordre élevé seront négligeables. On pourra donc tronquer  $p(x)$  pour obtenir :

$$\begin{aligned} f(x) &= \prod_{i=1}^p \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!} \cdot \left( 1 + \sum_{j < k} \rho_{jk} \frac{x_j - \lambda_j}{\sqrt{\lambda_j}} \frac{x_k - \lambda_k}{\sqrt{\lambda_k}} \right) \\ &= \prod_{i=1}^p \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!} \left( 1 + \sum_{j < k} \frac{\lambda_{jk}}{\sqrt{\lambda_j} \sqrt{\lambda_k}} \frac{x_j - \lambda_j}{\sqrt{\lambda_j}} \frac{x_k - \lambda_k}{\sqrt{\lambda_k}} \right) \end{aligned} \quad (25)$$

Une autre décomposition a été donnée par Krishnamoorthy [Kri51]

$$P(x_1, \dots, x_p) = \prod_{j=1}^p e^{-\lambda_j} \frac{\lambda_j^{x_j}}{x_j!} \exp \left( \sum_i \sum_j \theta_{ij} G(x_i) G(x_j) \dots \theta_{1\dots p} \prod_{i=1}^p G(x_i) \right)$$

où les  $G(\cdot)$  sont toujours les polynômes de Charlier.

## 5. Performances des classifieurs gaussiens et poissonniens

### 5.1 Données simulées

Le tableau suivant donne les résultats de reconnaissance sur des données poissonniennes bivariées avec des classifieurs gaussiens et des classifieurs poissonniens lorsque les paramètres sont petits (distributions mal approximées par une gaussienne). Le classifieur poissonnien est aussi bon ou meilleur que son équivalent gaussien.

Dans le cas multivarié, nous avons réalisé des simulations pour  $p = 8$ . Nous avons utilisé l'approximation de Bahadur pour les lois de Poisson. Outre la complexité de la génération de lois de Poisson dans  $N^8$ , il faut noter la lenteur de la reconnaissance dans le cas poissonnien par rapport à des lois gaussiennes sur des échantillons de taille 2 000.

En conclusion, le taux de reconnaissance avec le classifieur de Poisson est meilleur que celui obtenu avec le classifieur gaussien lorsque les paramètres des lois marginales sont inférieurs à 5, est pratiquement identique entre 5 et 12, est moins bon dès que ces valeurs de paramètres sont supérieures à 12.

## A PROPOS DE LA LOI DE POISSON MULTIVARIÉE

**Tableau 1.** Comparaison des performances des classifieurs poissonniens et gaussiens - données simulées

Numéro de simulation	Paramètres des échantillons		Matrices de confusion n = 2000				Pourcentage de bien classés n = 2000	
	$\pi_1$ : 1 <sup>er</sup> échantillon $\pi_2$ : 2 <sup>e</sup> échantillon	$\rho_{xy}$	classifieur gaussien		classifieur poissonnien		classifieur gaussien	classifieur poissonnien
1	$\pi_1$ : P (1.4 ; 3.3) $\pi_2$ : P (0.8 ; 1.1)	0.14 0.03	737 190	263 810	697 150	303 850	72.3	66.2
2	$\pi_1$ : P (2.2 ; 2.1) $\pi_2$ : P (1.6 ; 1.9)	0.51 0.45	648 405	352 595	888 564	112 436	62.10	66.2
3	$\pi_1$ : P (2.2 ; 4.1) $\pi_2$ : P (1.48 ; 1.7)	0.36 0.45	782 44	218 956	790 252	210 748	86.9	76.9
4	$\pi_1$ : P (1.19 ; 3.0) $\pi_2$ : P (0.7 ; 1.0)	# 0.0 # 0.0	746 175	254 825	756 181	244 819	78.5	78.7
5	$\pi_1$ : P (1.9 ; 2.9) $\pi_2$ : P (2.0 ; 1.9)	0.85 0.56	944 744	56 256	710 502	290 498	55.0	60.4
6	$\pi_1$ : P (6.0 ; 7.0) $\pi_2$ : P (5.0 ; 8.0)	0.61 0.47	868 55	132 945	897 112	103 888	90.6	89.2
7	$\pi_1$ : P (6.0 ; 7.0) $\pi_2$ : P (5.0 ; 8.0)	# 0 # 0	638 334	362 666	640 336	360 664	65.2	65.2
8	$\pi_1$ : P (5.0 ; 4.0) $\pi_2$ : P (4.0 ; 5.0)	0.22 0.22	824 319	176 681	853 147	147 853	75.2	85.3
9	$\pi_1$ : P (5.0 ; 4.0) $\pi_2$ : P (4.0 ; 5.0)	0.87 0.87	654 317	346 683	744 256	256 744	66.8	74.4

Note : la taille de  $\pi_1$  est égale à la taille de  $\pi_2 = 1\ 000$

### 5.2 Données de satellite Landsat

Nous avons utilisé le classifieur poissonnien sur des données provenant d'images de satellite. Il fallait classer 22 000 pixels, c'est-à-dire points de l'image satellite, caractérisés par 8 variables correspondant aux niveaux de gris dans 8 bandes de fréquence dans 53 classes définies *a priori*. Les niveaux de gris prenaient des valeurs discrètes entre 0 et 255. 3 400 points ont été utilisés pour l'apprentissage. Les valeurs moyennes des variables dans les différentes classes étaient comprises entre 50 et 100 et donc l'hypothèse poissonnienne multivariée ne convenait pas. Ceci explique les performances du classifieur poissonnien comparées à celles du classifieur gaussien par la règle du maximum de vraisemblance. Ce dernier reconnaît correctement 74 % des observations de l'ensemble test.

## 6. Conclusion

Nous avons fait une synthèse des propriétés des lois de Poisson multidimensionnelles. Lorsque le modèle sous-tendant les données est poissonnien, ce classifieur se comporte très bien et fait mieux dans de nombreux cas que son homologue normal. Il n'en reste pas moins que malgré les approximations et les développements en série, la manipulation d'une loi de Poisson multivariée reste calculatoirement lourde.

## BIBLIOGRAPHIE

- AITKEN A.C. (1944) *Statistical Mathematics*, Ed. Oliver and Boyd.
- BAHADUR P.R. (1961) "A Representation of the Joint Distribution of Response of N Dichotomous Items", *Studies of Item Analysis and Prediction*. H. Solomon ed. Stanford Un. Press, 158-168.
- CAMPBELL J.T. (1934) "The Poisson Correlation Function", *Proc. Edinburgh Math Soc.*, 4, 18-26.
- DWASS M. and TEICHER M. (1957) "On Infinitely Divisible Random Vectors", *The Ann. of Math. Stat.*, 28, 461-470.
- GULDBERG A. (1934) "On Discontinuous Frequency Functions of 2 Variables", *Skand. Aktuar.*, 17, 89-117.
- HOLGATE P. (1964) "Estimation for the Bivariate Poisson Distribution", *Biometrika*, 51, 241-245.
- KAWAMURA K. (1973) "The Structure of the Bivariate Poisson Distribution", *Kodai Math. Sem. Rep.*, 55, 246-256.
- KRISHNAMOORTHY A.S. (1951) "Multivariate Binomial and Poisson Distributions", *Sankhya*, 11, 117-124.
- MICLET L. (1984) *Méthodes structurelles pour la reconnaissance des formes*, Ed. Eyrolles.
- SAPORTA G. (1990) *Probabilités, Analyse de données et Statistique*, Ed. Technip.