REVUE DE STATISTIQUE APPLIQUÉE

MICHEL GOULARD

Mise en œuvre de l'algorithme EM pour l'estimation d'un modèle linéaire généralisé multinomial à effets aléatoires

Revue de statistique appliquée, tome 49, n° 4 (2001), p. 29-52

<http://www.numdam.org/item?id=RSA_2001__49_4_29_0>

Rev. Statistique Appliquée, 2001, **XLIX** (4), 29-52

MISE EN ŒUVRE DE L'ALGORITHME EM POUR L'ESTIMATION D'UN MODÈLE LINÉAIRE GÉNÉRALISÉ MULTINOMIAL À EFFETS ALÉATOIRES

Michel Goulard

Unité B.I.A., I.N.R.A., BP 27, 31326, CASTANET-TOLOSAN Cedex.

RÉSUMÉ

L'utilisation du modèle linéaire généralisé à effet aléatoire permet de bien résumer un jeu de données en gardant un modèle transposable à d'autres situations. Dans le cas multinomial nous étudions l'estimation de ce modèle par une méthode de type EM, en utilisant des méthodes MCMC pour les calculs d'espérance. Les éléments importants de la mise en œuvre de cette méthode sont le choix adaptif de la taille de l'échantillon simulé et l'utilisation intermédiaire de l'échantillonnage d'importance. Dans le cadre de l'analyse de données concernant le kiwi, nous montrons que le résumé fourni par le modèle à effet aléatoire est de meilleure qualité que celui obtenu grâce à un modèle à effets individuels fixes.

Mots-clés : *Modèle linéaire généralisé à effet aléatoire, algorithme EM, méthodes MCMC.*

ABSTRACT

Generalized linear mixed model gives a way to summarize a dataset keeping a transposable model. In a multinomial set-up we study estimation of the model using EM algorithm through MCMC methods for expectation calculus. Adaptative choice of sampling size and intermediate use of importance sampling are important points of the implementation of the method. In the analysis of a kiwi tree dataset, we exhibit a better summary produced by mixed model than those obtained by models with fixed effects.

Keywords : *Generalized linear mixed model, EM algorithm, MCMC methods.*

1. Introduction

Le modèle linéaire généralisé multinomial (Fahrmeir & Tutz 1994) est un modèle simple pour décrire la distribution d'une variable discrète à plusieurs modalités en fonction de variables explicatives. L'estimation de paramètres dans le cadre du modèle linéaire généralisé avec effets aléatoires a fait l'objet d'un certain nombre d'études assez récentes. L'ensemble des méthodes développées dans ces études font un usage intensif de l'ordinateur, et toutes utilisent une méthode de type MCMC. Certains travaux relèvent du paradigme bayésien (Gamerman 1997, Zeger & Karim 1991). L'utilisation des idées de l'échantillonnage d'importance pour calculer la vraisemblance est une autre voie (Geyer & Thompson 1992). La méthode de type EM, que nous avons choisie d'utiliser, appartient à la classe qui considère

le problème comme étant à données manquantes (McCulloch 1997, Booth & Hobert 1999, Quintana, Liu & Pino 1999, Tanner 1993). L'objectif du travail était alors d'effectuer une mise en œuvre de la méthode EM pour le modèle multinomial, sur des données réelles, en prenant en compte un certain nombre de considérations tirées de la bibliographie, comme l'adaptativité de la taille de l'échantillon simulé (Booth & Hobert 1999) et l'utilisation de l'échantillonnage d'importance (Quintana, Liu & Pino 1999). L'utilisation simultanée de MCMC et de l'échantillonnage d'importance est rencontrée dans la littérature pour le filtrage de systèmes dynamiques quand on considère une stratégie bayésienne, la méthode est connue sous le nom de « particle filters » (voir Doucet 1998, Liu & Chen 1998 et les références contenues dans ces articles).

Dans le cadre d'une analyse statistique portant sur le kiwi, nous avons utilisé le modèle multinomial pour expliquer la distribution de calibres sur une liane en fonction du nombre de cannes portées par la liane et du nombre de fleurs. Les données, dont nous disposions, décrivaient pour un certain nombre de lianes la répartition des kiwis en classe de calibre. Ces données sont issues d'études menées sur la conduite de production du kiwi (Agostini 1995). L'utilisation du modèle linéaire généralisé multinomial a donné un résultat qui n'était pas satisfaisant au niveau de chaque liane. Un modèle multinomial, prenant en compte l'effet fixe liane, donne un meilleur résultat au niveau individuel. Cependant ce modèle ne permet pas de généraliser à d'autres situations et l'adéquation du modèle est rejetée. Le modèle linéaire généralisé multinomial avec effets aléatoires nous a fourni un meilleur résumé des données.

L'article décrit la méthode d'estimation rassemblant les ajouts et les remarques issus de la bibliographie, puis montre comment sa mise en œuvre effective sur des données permet de mieux « coller » aux données en gardant un modèle tranposable à d'autres situations.

Dans une première partie nous présentons le modèle linéaire généralisé multinomial qui sert de base au travail. Le modèle à effets aléatoires ainsi que la méthode d'estimation sont présentés dans une deuxième partie. La mise en œuvre de la méthode sur les données issues d'une expérimentation portant sur le kiwi est présentée dans une troisième partie. Les fonctions permettant l'estimation des différents modèles décrits dans l'article, ont été écrites en S et en Fortran.

2. Le modèle linéaire généralisé multinomial.

Soit Y une variable aléatoire discrète, prenant les valeurs $1, \ldots, J$, qui définit une affectation parmi J classes. Par souci de simplification nous allons dans la présentation nous restreindre au cas d'une variable explicative.

Soit X une variable explicative pour Y. Un modèle simple pour décrire la probabilité que Y prenne la valeur j sachant $X = x$, X étant à valeurs discrètes ou dans $\mathbb{R}$, est :

$$P(Y = j|X = x) = \frac{\exp(\alpha_j + \beta_j X)}{\sum_l \exp(\alpha_l + \beta_l X)} \quad (*)$$

Une contrainte sur les α_j et β_j est nécessaire pour rendre ce modèle identifiable. Par exemple on prend $\alpha_J = \beta_J = 0$. Ce modèle est associé à un principe de maximisation d'une utilité (Fahrmeir & Tutz 1994, 70-71).

Supposons maintenant que l'on observe $(n_{i,1}, \ldots, n_{i,J}, x_i)$ pour $i = 1, \ldots, n$ où $n_{i,j}$ est le nombre de réalisations pour lesquelles $Y = j$ dans des expériences indépendantes. L'estimation des paramètres du modèle $(*)$ se fait par maximisation de la vraisemblance qui s'écrit :

$$f(\alpha, \beta) = \prod_{i=1}^{n} \prod_{j=1}^{J} P(Y = j | X = x_i)^{n_{i,j}}.$$

En passant au logarithme on est conduit à maximiser

$$L(\alpha, \beta) = \sum_{i=1}^{n} \left(\sum_{j=1}^{J-1} \{(\alpha_j + \beta_j x_i) n_{i,j}\} - n_{i,+} \log\,(1 + \sum_{j=1}^{J-1} \exp\,(\alpha_j + \beta_j x_i)) \right)$$

où $n_{i,+} = \sum_j n_{i,j}$ et où α (resp. β) est le vecteur des α_j (resp. β_j) $(1 \leqslant j \leqslant J-1)$.

La méthode d'estimation possède les propriétés générales liées à la méthode du maximum de vraisemblance (Fahrmeir & Tutz 1994). En particulier la loi asymptotique de l'estimateur de $\theta = (\alpha, \beta)$ est gaussienne de matrice de variance-covariance l'inverse de la matrice d'information de Fisher. Le test d'hypothèse linéaire sur les paramètres peut se faire en utilisant le rapport de vraisemblance et on dispose d'une statistique d'adéquation globale du modèle qui est la déviance :

$$D(\alpha, \beta) = 2(L(\alpha, \beta) - \sum_{i=1}^{n} n_{i,j} \log(n_{i,j}/n_{i,+})).$$

L'adéquation individuelle du modèle peut être étudiée au travers des résidus de Pearson définis par :

$$r_i = V^{-1/2}(\hat{\theta})(n_i - \hat{\mu}_i)$$

où $\hat{\theta}$ est l'estimateur des paramètres, n_i est le vecteur des comptage, $V(\hat{\theta})$ sa matrice de variance-covariance et $\hat{\mu}_i$ est l'espérance de n_i calculée sous $\theta = \hat{\theta}$.

Dans certains cas on peut utiliser un modèle, plus complet que $(*)$, dans lequel un effet individuel est présent :

$$P_{i,j} = \frac{\exp\,(\alpha_j + \beta_j \gamma_i)}{\sum_l \exp\,(\alpha_l + \beta_l \gamma_i)} (**)$$

avec comme contraintes pour identifier le modèle α_J et β_J pris égaux à 0, et deux valeurs des effets individuels γ_i fixés. Cette dernière contrainte est prise car le modèle $(**)$ est invariant par transformation linéaire des coefficients de la forme :

$\gamma_i \to a\gamma_i + b$, $\alpha_j \to \alpha_j - b\beta_j$ et $\beta_j \to \beta_j/a$. L'estimation de ce modèle peut être faite par maximum de vraisemblance comme le modèle simple et ce qui a été présenté précédemment reste valable. Ce modèle prend en compte une variabilité individuelle des comportements mais ne permet pas de transposer à une autre situation par exemple par une prédiction. D'autre part le nombre de paramètres à estimer est évidemment pénalisant. On préfère souvent le modèle à effets aléatoires présenté dans la partie suivante.

3. Le modèle à effets aléatoires.

Nous considèrons maintenant le modèle :

$$P(Y = j|X = x_i) = \frac{\exp(\alpha_j + \beta_j x_i + u_{i,j})}{\sum_l \exp(\alpha_l + \beta_l x + u_{i,l})} \text{ pour } j = 1, \ldots, J-1$$

$$P(Y = J|X = x_i) = \frac{1}{\sum_l \exp(\alpha_l + \beta_l x + u_{i,l})}.$$

où les composantes $u_{i,j}$ sont des réalisations de variables aléatoires U_j pour $j = 1, \ldots, J-1$ gaussiennes centrées et de matrice de variance covariance Σ. Ce modèle prend en compte la variabilité individuelle en gardant une certaine généralité. Il est possible de faire de la prédiction pour une nouvelle liane et le nombre de paramètres n'est pas trop important.

L'estimation de ce modèle n'est pas directe, les $u_{i,j}$ étant non-observées. Certains auteurs utilisent la méthodologie bayésienne pour résoudre le problème (Gamerman 1997, Zeger & Karim 1991). Cela pose le problème du choix des lois *a priori*. La difficulté de ce choix dans notre cas nous a conduit à ne pas considérer cette approche qui donc ne sera pas présentée. La vraisemblance du modèle s'écrit :

$$g(\alpha, \beta, \Sigma) = \prod_{i=1}^{n} \prod_{j=1}^{J} \int P(Y = j|X = x_i)^{n_{i,j}} \Phi(u_{i,j}|\Sigma) du_{i,j}$$

$$= \prod_{i=1}^{n} \int h(n_i|x_i, U, \alpha, \beta) \Phi(U|\Sigma) dU$$

où Φ désigne la densité gaussienne utilisée pour U et $h(n_i|x_i, U, \alpha, \beta)$ est la densité de la loi des comptages pour l'individu i sachant x_i, U, α, β. L'évaluation analytique de cette quantité n'est pas possible.

L'évaluation de la vraisemblance par Monte-Carlo comme suggérée dans Geyer & Thompson (1992) est une possibilité. On simule des $u_{i,l}$ pour $l = 1, \ldots, M$ suivant une $N(O, \Sigma^0)$, Σ^0 étant une valeur de départ que l'on se donne, puis on maximise $g(\alpha, \beta, \Sigma)$ à partir de son évaluation par échantillonnage d'importance en utilisant l'échantillon $u_{i,l}$. Ce qui donne des valeurs $\hat{\alpha}$, $\hat{\beta}$ et $\hat{\Sigma}$. L'étape suivante consiste à

remplacer Σ^0 par $\Sigma^1 = \hat{\Sigma}$ et à itérer le processus. Cette méthode ne converge pas très vite en général et donc n'a pas été retenue.

3.1. Méthode d'estimation choisie

La méthodologie développée par McCulloch (1997) est basée sur un algorithme EM avec Monte-Carlo : la question d'estimation est considérée comme étant un problème à données manquantes. On note dans la suite $E(V(U)|N)$ l'espérance conditionnelle de la fonction V prise sur les U conditionnellement aux observations, la fonction V pouvant dépendre des observations et des paramètres. L'algorithme de type EM pour estimer le modèle est alors (McCulloch 1997) :

- on choisit des valeurs $\alpha^{(0)}$, $\beta^{(0)}$ et $\Sigma^{(0)}$ de départ, et l'indice de boucle $m = 0$
- on calcule :
  - $\alpha^{(m)}$ et $\beta^{(m)}$ qui maximisent $\sum_{i=1}^{n} E(\log h(n_i|x_i, U)|N)$
  - $\Sigma^{(m)}$ qui maximise $E(\log \Phi(U|\Sigma)|N)$
  - $m = m + 1$
- Si le processus converge les valeurs atteintes sont les estimateurs de α, β et Σ, sinon on itère le calcul.

Le critère d'arrêt porte sur la variation relative des paramètres $\alpha^{(m)}$ et $\beta^{(m)}$. La difficulté réside dans le calcul de l'espérance sous la loi *a posteriori* des U sachant les données. Pour cela on utilise soit une méthode de type Metropolis (Robert 1996), soit l'échantillonnage d'importance (Robert 1996, Tanner 1993).

En pratique nous avons travaillé sous le modèle équivalent où α n'apparaît pas dans l'écriture des $P(Y = j|X = x)$ mais pour lequel U suit une gaussienne de moyenne α et de variance Σ. Cela accélère les convergences, en particulier la maximisation de la vraisemblance ne porte que sur β, α étant obtenu par une simple moyenne empirique. Par la suite nous décrivons la procédure Metropolis et celle d'échantillonnage d'importance dans ce modèle équivalent. Il faut noter qu'au niveau de l'algorithme cette modification entraine que le critère d'arrêt ne portera plus que sur la variation relative des paramètres $\beta^{(m)}$.

3.2. Procédure type Metropolis

La procédure Metropolis permet de simuler un vecteur $u = (u_1, \ldots, u_{J-1})$ suivant la loi *a posteriori* sachant les données pour chaque i. Cette loi est donnée par :

$$\pi_i(u|\alpha, \beta, \Sigma) \propto \prod_{j}^{J} P_{i,j}(u|\beta)^{n_{i,j}} \Phi((u_1, \ldots, u_{J-1})|\alpha, \Sigma)$$

où $P_{i,j}(u|\beta) = P(Y = j|X = x_i)$. On utilise une procédure itérative pour réaliser la simulation. Notons qu'ici on n'utilise pas la méthode la plus générale (Robert 1996)

mais celle avec tirage indépendant. À l'étape 0, on initialise $U^{(0)}$ par un tirage à partir d'une loi ρ. Soit $U^{(l)}$ la valeur de U obtenue à l'étape l, la valeur $U^{(l+1)}$ est obtenue par la règle :

$$U^{(l+1)} = \begin{cases} U^* & \text{avec la probabilité } p_M = \min\left(1, \dfrac{\pi(U^*|\alpha,\beta,\Sigma)\rho(U^{(l)})}{\pi(U^{(l)}|\alpha,\beta,\Sigma)\rho(U^*)}\right) \\ U & \text{avec la probabilité } 1 - p_M \end{cases}$$

où U^* est obtenue par un tirage suivant ρ. Au bout d'un nombre d'itérations suffisamment grand, on obtient un vecteur U qui suit approximativement la loi demandée. Le choix de la loi ρ est évidemment important. Si on prend pour ρ la loi gaussienne $\Phi(\alpha, \Sigma)$ la probabilité p_M s'écrit :

$$p_M = \min(1, \prod_j^J P_{i,j}(U^*|\beta)^{n_{i,j}} / P_{i,j}(U^{(l)}|\beta)^{n_{i,j}})$$

les quantités $\rho(U^{(l)})$ et $\rho(U^*)$ ayant disparu par simplification du fait du choix de ρ.

Supposons maintenant que nous ayons à notre disposition un échantillon $(u_{i,1}, \ldots, u_{i,M})$, pour $i = 1, \ldots, n$, l'étape de calcul devient :

$$\beta^{(m)} = \operatorname{Arg}\min_\beta \sum_{i=1}^n \frac{1}{M} \sum_{s=1}^M \sum_{j=1}^J n_{i,j} \log P_{i,j}(u_{i,s}|\beta)$$

$$\alpha^{(m)} = \frac{1}{nM} \sum_{i=1}^n \sum_{s=1}^M u_{i,s} \text{ et}$$

$$\Sigma^{(m)} = \left(\frac{1}{nM} \sum_{i=1}^n \sum_{s=1}^M u_{i,s}^T u_{i,s}\right) - \alpha^{(m)T} \alpha^{(m)}$$

où v^T désigne le vecteur transposé de v.

3.3. Procédure type échantillonnage d'importance

Dans le cas de l'échantillonnage d'importance, on engendre pour chaque i un M-échantillon $(u_{i,1}, \ldots, u_{i,M})$ de vecteurs suivant une loi ρ. L'espérance $E_{\pi_i}(V(U))$, pour une fonction V quelconque, est estimée par :

$$\frac{\sum_1^M w_{i,s} V(u_{i,s})}{\sum_1^M w_{i,s}}$$

avec $w_{i,s} = \pi_i(u_{i,s}|\alpha, \beta, \Sigma)/\rho(u_{i,s})$. Là encore le choix de la loi ρ est déterminant car il conditionne le nombre de valeurs « efficaces » c'est-à-dire qui ont un poids non-négligeable dans le calcul. Pour notre cas l'étape de calcul devient alors :

$$\beta^{(m)} = \operatorname*{Arg\,min}_{\beta} \sum_{i=1}^{n} \frac{1}{\sum_{1}^{M} w_{i,s}} \sum_{s=1}^{M} w_{i,s} \sum_{j=1}^{J} n_{i,j} \log P_{i,j}(u_{i,s}|\beta)$$

$$\alpha^{(m)} = \frac{1}{n \sum_{1}^{M} w_{i,s}} \sum_{i=1}^{n} \sum_{s=1}^{M} w_{i,s} u_{i,s} \text{ et}$$

$$\Sigma^{(m)} = \left(\frac{1}{n \sum_{1}^{M} w_{i,s}} \sum_{i=1}^{n} \sum_{s=1}^{M} w_{i,s} u_{i,s}^T u_{i,s} \right) - \alpha^{(m)T} \alpha^{(m)}.$$

En pratique, comme suggéré par Quintana, Liu & Pino (1999), on peut n'utiliser l'échantillonnage d'importance qu'entre deux étapes de Metropolis. Si pour une itération donnée on a engendré $(u_{i,1}, \ldots, u_{i,M})$ suivant la loi $\pi_i(U|\alpha, \beta, \Sigma)$ pour $i = 1, \ldots, n$ par Metropolis, on utilisera ces vecteurs pour les itérations suivantes, dans la procédure d'échantillonnage d'importance, pour les paramètres $(\alpha', \beta', \Sigma')$ qui ne sont pas « trop loin » de (α, β, Σ). On juge de cet éloignement en estimant le nombre moyen de valeurs $u_{i,m}$ efficaces (qui interviennent dans le calcul) par $\frac{1}{n} \sum_{i=1}^{n} \sum_{1}^{M} w_{i,s}$, sachant que si tous les poids valent 1 cette quantité vaut M. Une autre estimation de ce nombre efficace peut-être fait en utilisant le coefficient de variation (Liu & Chen 1995).

Dès que $(\alpha', \beta', \Sigma')$ et (α, β, Σ) sont éloignés (c.a.d. le nombre efficace devient inférieur, par exemple, à 0.8 M) on revient à une procédure Metropolis pour engendrer un échantillon suivant $\pi_i(U|\alpha', \beta', \Sigma')$. L'utilisation intermédiaire de l'échantillonnage d'importance permet évidemment d'accélérer le calcul.

3.4. Taille de l'échantillon Monte-Carlo

Un autre point important à prendre en compte est la taille de l'échantillon simulé : lorsque l'algorithme n'est pas près de la solution stationnaire il n'est pas nécessaire d'avoir recours à une taille d'échantillon très importante. Par contre il faut augmenter cette taille au fur et à mesure de l'évolution de l'algorithme, en effet elle conditionne la stabilité de l'estimation de l'espérance. On veut donc modifier la taille en fonction du raffinement que l'on veut obtenir. Pour cela on utilise la règle d'augmentation suggérée dans Booth & Hobert (1999) : la taille de l'échantillon est

augmentée chaque fois que la variation en norme des paramètres estimés n'est pas significative au sens d'un test utilisant la matrice de variance-covariance théorique (asymptotique) de l'estimateur du paramètre obtenu à cette étape. Notons que dans notre mise en œuvre l'augmentation de la taille de l'échantillon est couplée à une utilisation de la méthode Metropolis.

3.5. Remarque

Dans le modèle précédent on suppose qu'il y a $J-1$ composantes qui agissent, chacune étant spécifique de la classe à laquelle elle est associée. Leur action peut-être corrélée, cas général, ou indépendante, cas Σ diagonale. La décomposition spectrale de la matrice de corrélation (estimée) associée peut montrer que ces composantes sont sur un espace de dimension réduit. L'utilisation du modèle précédent en considérant que Σ n'est pas de rang $J-1$, n'est pas correcte du point de vue théorique. On peut aussi introduire le modèle :

$$P(Y=j|X=x_i)=\frac{\exp\left(\alpha_j+\beta_j x_i+\sum_1^K \gamma_{j,k}u_{i,j,k}\right)}{\sum_l \exp\left(\alpha_l+\beta_l x+\sum_1^K \gamma_{l,k}u_{i,l,k}\right)}$$

où les $u_{i,l,k}$ sont des réalisations indépendantes de variables U_k, $k=1,\ldots,K$ gaussiennes centrées réduites. Le nombre K de facteurs introduits est inférieur à $J-1$ et $\alpha_J=\beta_J=\gamma_{J,1}=\ldots=\gamma_{J,K}=0$. L'estimation de α, β et γ peut se faire en adaptant la procédure EM décrite précédemment.

4. Utilisation pour les données kiwi.

Le kiwi (Actinidia deliciosa cv. Hayward) est une liane dioïque non greffée dont la culture mono-variétale est soumise à un mode de conduite dominant (le T-bar), et dont la principale étape limitant le rendement est l'élaboration du nombre de fruits. C'est en outre une culture qui ne subit pas de traitement phytosanitaire. L'ensemble de ces caractéristiques confère au kiwi un statut de culture « modèle » pour l'étude des interventions techniques en arboriculture fruitière qui sont peu nombreuses chez cette espèce : irrigation, fertilisation, taille et éclaircissage. En assurant une alimentation hydrominérale non limitante, mais raisonnable, on peut même réduire la culture du kiwi aux deux actes techniques que sont la taille d'hiver et l'éclaircissage des jeunes fruits au printemps. L'objectif de l'analyse était de construire, et d'estimer, un modèle statistique simple qui doit décrire la distribution des calibres en fonction de la charge en fruits et/ou de la charge en rameaux. A partir de ce modèle on peut, en prenant en compte une structure de prix, évaluer économiquement des pratiques de taille et/ou d'éclaircissage.

Nous disposions d'un jeu de données pour 71 lianes de kiwi en 1995 pour lesquelles la charge en fruits (nombre de fruits sur la liane y compris les fruits éclaircis), le nombre de cannes sur la liane et la répartition dans 11 classes de calibres avaient été notés.

La figure 1 présente pour les 71 lianes et les 11 classes de calibre, les fréquences observées en fonction du nombre de cannes sur la liane. On remarque un certain nombre de points extrêmes, pour lesquels certaines fréquences sont très différentes des autres, qui correspondent à une dizaine de lianes. Ces lianes seront omises pour les ajustements suivants d'une part car leur fonctionnement parait singulier et donc n'est pas intéressant pour un résumé, d'autre part leur prise en compte pourrait gêner l'estimation des paramètres, en particulier les convergences de l'algorithme EM. Dans la suite nous conservons 61 lianes pour les calculs.

La modélisation de ces fréquences dans les classes de calibre a été faite en utilisant le modèle linéaire généralisé multinomial. Les tableaux 1 et 2 donnent les paramètres estimés, avec les écarts-types théorique des estimateurs. Le tableau 3 donne les valeurs de déviance qui permettent de faire des tests d'ajout des variables ou de juger de l'adéquation des modèles. Si globalement l'ajout des deux variables est significatif, on pourrait certainement raffiner le modèle en ne prenant que la variable nombre de cannes pour certaines classes.

TABLEAU 1

Valeurs des coefficients α estimés par maximum de vraisemblance et écarts-types d'estimation, dans le cadre du modèle simple.

coefficient	Valeur	Écart-type
α_1	-1,99	0,17
α_2	-1,90	0,18
α_3	-1,40	0,18
α_4	-0,99	0,18
α_5	-0,70	0,18
α_6	-0,08	0,18
α_7	-0,56	0,19
α_8	1,11	0,20
α_9	0,94	0,22
α_{10}	-0,56	0,23

L'ajustement de ce modèle, avec la charge en fruit et le nombre de cannes comme variables explicatives, semble structurellement correct. Cependant pour certaines lianes l'ajustement est assez mauvais (figure 2), on observe des résidus trop élevés (figures 3 et 4) et le test d'adéquation utilisant la déviance (tableau 3) conduit à un rejet.

TABLEAU 2

Valeurs des coefficients β estimés par maximum de vraisemblance, et écarts-types d'estimation entre parenthèses, dans le cadre du modèle simple et du modèle à effet aléatoire. Les variances pour l'estimation dans le cadre du modèle à effet aléatoire sont obtenues en utilisant la formule de Louis (1982) pour la matrice d'information.

coefficient	Simple	avec Effet aléatoire
β_1^1	2,59 (0,30)	2,33 (0,22)
β_2^1	2,65 (0,30)	2,47 (0,22)
β_3^1	2,52 (0,31)	2,34 (0,21)
β_4^1	2,76 (0,31)	2,58 (0,21)
β_5^1	2,84 (0,31)	2,62 (0,21)
β_6^1	2,73 (0,31)	2,59 (0,20)
β_7^1	2,56 (0,32)	2,44 (0,20)
β_8^1	2,01 (0,34)	1,87 (0,20)
β_9^1	0,89 (0,39)	0,80 (0,22)
β_{10}^1	0,82 (0,40)	0,71 (0,23)
β_1^2	1,40 (0,28)	1,19 (0,19)
β_2^2	0,99 (0,28)	0,72 (0,20)
β_3^2	0,89 (0,28)	0,58 (0,19)
β_4^2	0,43 (0,28)	0,14 (0,19)
β_5^2	0,21 (0,28)	-0,09 (0,18)
β_6^2	0,02 (0,29)	-0,30 (0,18)
β_7^2	-0,29 (0,29)	-0,66 (0,18)
β_8^2	-0,56 (0,32)	-1,03 (0,18)
β_9^2	-0,49 (0,36)	-0,91 (0,20)
β_{10}^2	-0,60 (0,37)	-0,80 (0,21)

TABLEAU 3

*Tableau donnant pour les différents modèles utilisés la valeur de la déviance. Les nombres de ddl résiduel sont obtenus par la différence du nombre d'observations (671 = 61*11) et du nombre de paramètres du modèle.*

Modèle utilisé	nombre de paramètres	Déviance	nombre de ddl résiduel
Simple sans variable	10	9896,8	661
Simple avec nbre de fruits	20	9753,4	651
Simple avec nbre de cannes	20	9626,2	651
Simple avec 2 variables	30	6665,6	641
avec Effet Fixe Liane	69	1709,4	602
avec Effet Aléatoire	75	670,6	596

Pour remédier à la relative faiblesse du modèle, nous introduisons un effet fixe liane et donc nous utilisons le modèle $(**)$. La qualité de l'ajustement est globalement améliorée du point de vue des résidus (figures 5 et 6), la reconstruction par liane est satisfaisante (figure 2) mais on a toujours un test d'adéquation qui rejette le modèle (tableau 3).

Le modèle $(**)$ estimé est tout juste satisfaisant et de plus gênant pour atteindre une certaine généralité. D'autre part les variables explicatives n'interviennent plus.

On peut dans un premier temps les réintroduire en faisant une régression des effets lianes sur les variables considérées. Une autre façon de faire est d'utiliser le modèle à effets aléatoires qui a l'avantage d'avoir moins de paramètres à estimer.

La méthode d'estimation présentée précédemment a été utilisée pour estimer les divers paramètres. La taille de l'échantillon des vecteurs aléatoires simulés est fixée à $M = 100$ au départ. A la convergence elle est de $M = 3122$. Pour ce qui concerne la méthode Metropolis, le nombre d'itérations a été fixé à 2000, pour obtenir chacune des M valeurs de vecteur que l'on considère comme issu de la distribution cible. Ce nombre d'itérations (2000) est issu d'une étude préalable évaluant la stabilité de la distribution empirique obtenue après un nombre d'itérations donné. Comme attendue la méthode Metropolis est très lourde en temps calcul, avec 2000 * M évaluations de « vraisemblance », et l'utilisation de l'échantillonnage d'importance accélère les calculs.

Pour ne pas alourdir l'article, nous ne présentons dans la suite que des résultats sur les termes diagonaux $\sigma_k^2 = \Sigma_{kk}$ de la matrice Σ.

Sur la figure 8 sont représentées l'évolution de certains paramètres au cours des itérations de l'algorithme. La fluctuation des paramètres β devient très faible à partir d'une centaine d'itérations. La convergence aurait été obtenue après 113 itérations,

si le critère d'arrêt avait été une variation relative des paramètres β inférieure à 10^{-3}. Les paramètres de variance sont trés rapidement peu fluctuants aprés une vingtaine d'itérations. Les paramètres de moyenne α fluctuent plus longtemps, leurs fluctuations étant liées à la convergence des paramètres β. Les tableaux 2 et 4 donnent les estimations obtenues aprés 120 itérations, avec les écarts-type théoriques.

TABLEAU 4

Valeurs des coefficients α et des écarts-types estimés dans le cadre du modèle à effet aléatoire et écarts-types d'estimation entre parenthèses. Les écarts-types d'estimation sont calculés à partir de la formule empirique de l'algorithme.

coefficient	Valeur
α_1	-1,3 (0,028)
α_2	-1,15 (0,025)
α_3	-0,58 (0,024)
α_4	-0,18 (0,023)
α_5	0,17 (0,022)
α_6	0,91 (0,020)
α_7	1,41 (0,018)
α_8	2,03 (0,014)
α_9	1,63(0,010)
α_{10}	0,94 (0,007)
σ_1	1,57 (0,062)
σ_2	1,41 (0,051)
σ_3	1,32 (0,044)
σ_4	1,28 (0,042)
σ_5	1,23 (0,038)
σ_6	1,15 (0,033)
σ_7	1,01 (0,026)
σ_8	0,80 (0,016)
σ_9	0,54 (0,007)
σ_{10}	0,38 (0,004)

Le test d'adéquation du modèle conduit à ne pas le rejeter (Tableau 3). La décomposition en coordonnées principales de la matrice de corrélation des composantes aléatoires (figure 9) montre que l'on pourrait réduire le nombre de composantes et faire une estimation dans un cadre avec deux composantes aléatoires (*cf.* Remarque 3.5). L'une, principale (87% de 'inertie) exprime une différence globale portant sur les classes intermédiaires, l'autre (11% de l'inertie) montre un différentiel sur les classes extrêmes.

Une estimation des composantes aléatoires a été effectuée en utilisant la loi *a posteriori* avec les paramètres estimés. On effectue une estimation par moyenne a posteriori en utilisant le procédé MCMC et en faisant la moyenne des composantes par liane. Le calcul des résidus, quand on intègre l'estimation des composantes avec celle du modèle, montre qu'on améliore la qualité de la reconstruction des fréquences (figure 2 et 7). Sur la figure 10 on peut voir une estimation par intervalle des probabilités individuelles en utilisant la loi *a posteriori*. Si globalement cela donne des résultats satisfaisants, cela met quand même en évidence des problèmes individuels pour certaines lianes. Certains décalages peuvent être liés à la qualité de la convergence individuelle après 2000 itérations de la méthode MCMC pour obtenir la loi *a posteriori*.

5. Conclusions

Ce travail montre l'intérêt du modèle généralisé multinomial avec effets aléatoires. Cela permet d'améliorer la qualité d'ajustement pour tenir compte des variabilités individuelles en gardant un modèle qui n'est pas anecdotique. L'estimation dans le cadre de ce modèle est assez délicate et surtout assez lourde en temps calcul. La mise à jour de la taille de l'échantillon de travail et l'utilisation intermédiaire de l'échantillonnage d'importance sont déterminantes pour réduire ce temps calcul.

6. Remerciements

Je remercie Dominique Agostini (Station de Recherches Agronomiques de Corse, INRA-CIRAD, San Giuliano) de m'avoir fourni le jeu de données kiwi.

Références

AGOSTINI D., (1995) *La floraison du kiwi (Actinidia deliciosaa cv. Hayward) : analyse de la variabilité et simulation par un modèle stochastique.* Thèse de doctorat, Université de Lyon I.

BOOTH J. G., & Hobert J. P., (1999) Maximizing generalized linear mixed model likelihood with an automated Monte Carlo algorithm. *J. Roy. Statist. Soc. Ser. B*, 61, 1, 265-285.

DOUCET A., (1998) On sequential simulation-based methods for Bayesian filtering. Technical report CUED/F-INFENG/TR.310.

FAHRMEIR L. & TUTZ G., (1994) *Multivariate statistical modelling based on generalized linear models.* Springer-Verlag, New York, 425p.

GAMMERMAN D., (1997) Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7, 57-68.

GEYER C. J. & THOMPSON E. A. (1992) Constrained Monte Carlo maximum likelihood for dependent data. *J. Roy. Statist. Soc. Ser. B* 54, 3, 657-699.

LIU J. S. & CHEN R. (1995). Blind deconvolution via sequential imputations. *J.A.S.A.*, 90, 430, 567-576.

LIU J. S. & CHEN R. (1998) Sequential Monte Carlo methods for dynamic systems. *J.A.S.A.*, 93, 443, 1032-1044.

LOUIS T. A. (1982) Finding the observed information matrix when using the *EM* algorithm. *J. Roy. Statist. Soc. Ser. B* 44, 2, 226-233.

McCULLOCH C. E. , (1997) Maximum likelihood algorithms for generalized linear mixed models. *JASA*, 92, 437, 162-170.

QUINTANA F. A., LIU J. S. & DEL PINO G. E., (1999) Monte-Carlo EM with importance reweighting and its applications in random effects models. *Computational Statistics & Data Analysis*, 29, 429-444.

ROBERT C., (1996) *Méthodes de Monte-Carlo par chaînes de Markov.* Statistique mathématique et probabilité, Economica, Paris, 340 p.

TANNER M. A, (1993) *Tools for statistical inference.* 3^{rd} *edition.* Springer, New-York, 207 p.

ZEGER S. L. & M. REZAUL Karim, (1991) Generalized linear models with random effects; A Gibbs sampling approach. *JASA*, 86, 413,79-86.

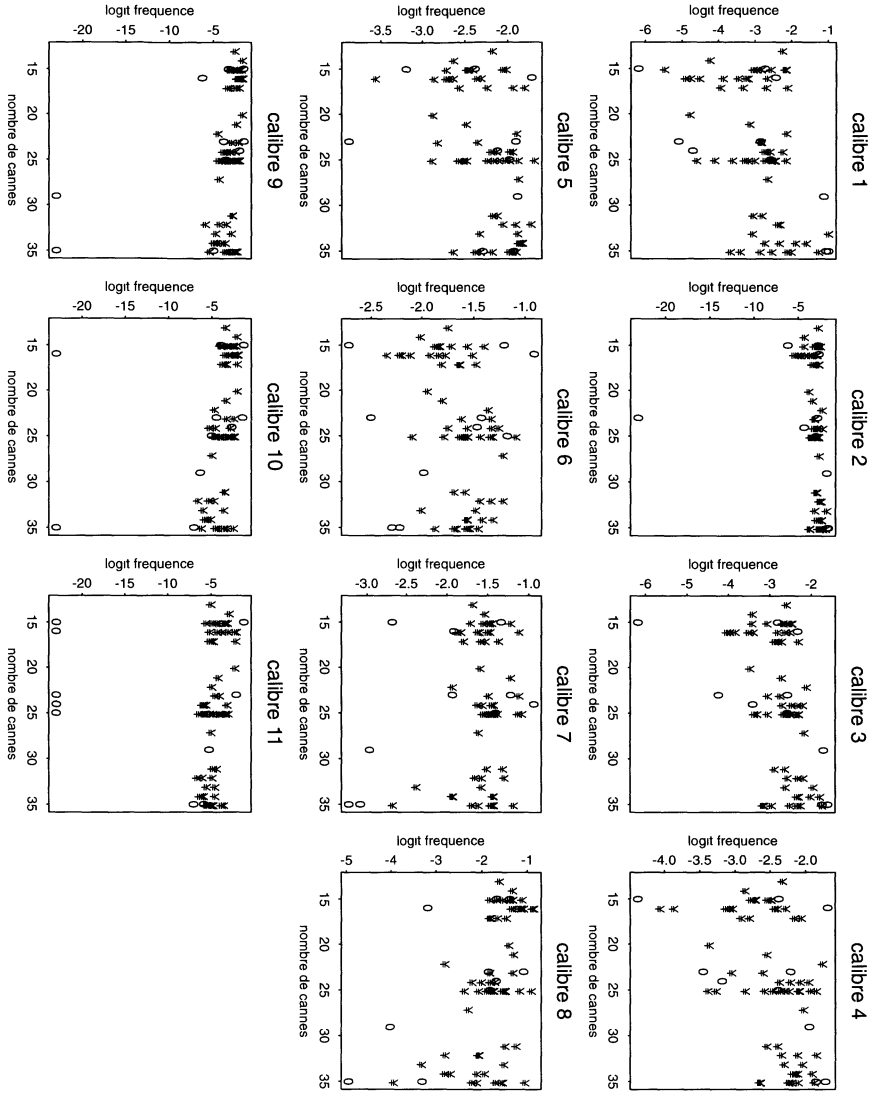

FIGURE 1

Graphique des logit de fréquence dans chaque classe de calibre en fonction du nombre de cannes. Les points extrêmes sont repérés par des 0.

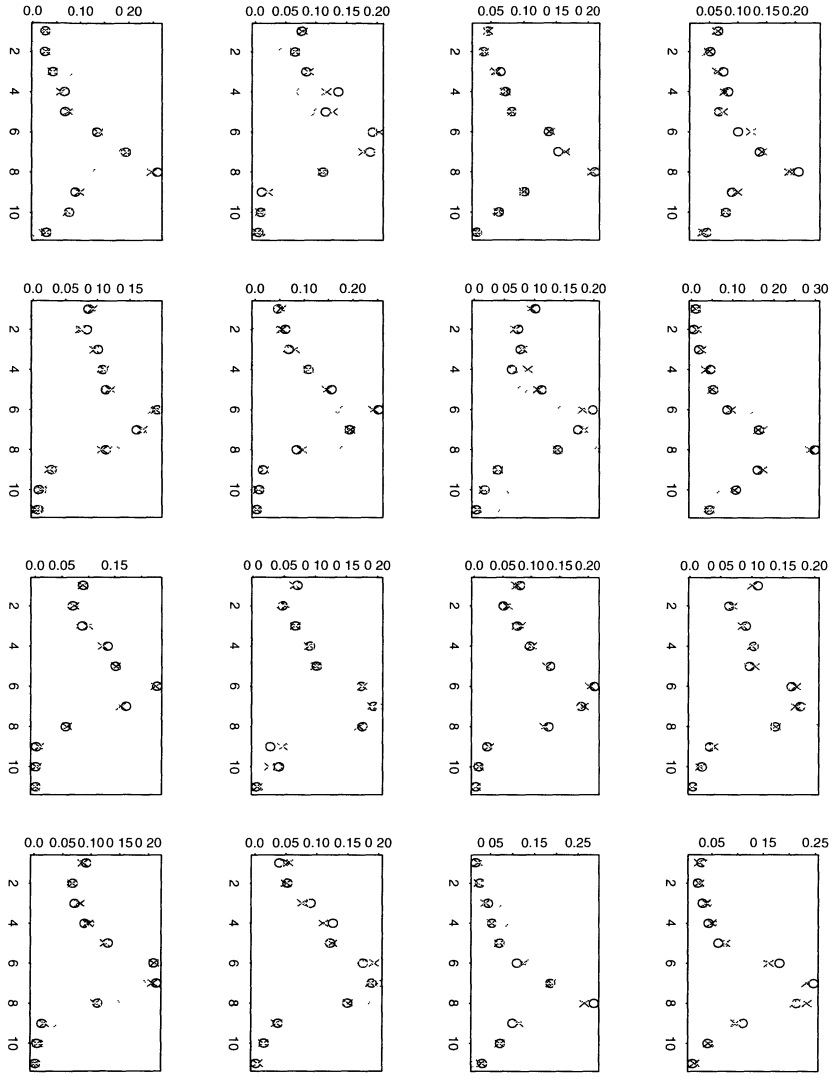

FIGURE 2

Représentation pour 16 lianes des fréquences observées et des fréquences ajustées par le modèle multinomial. Le numéro de la classe de calibre est en abscisse et les fréquences sont en ordonnées : « 0 » fréquence observée, « S » valeur estimée par le modèle multinomial simple, « T » valeur estimée par le modèle multinomial avec effet liane, «X» valeur estimée par le modèle multinomial avec effet aléatoire.

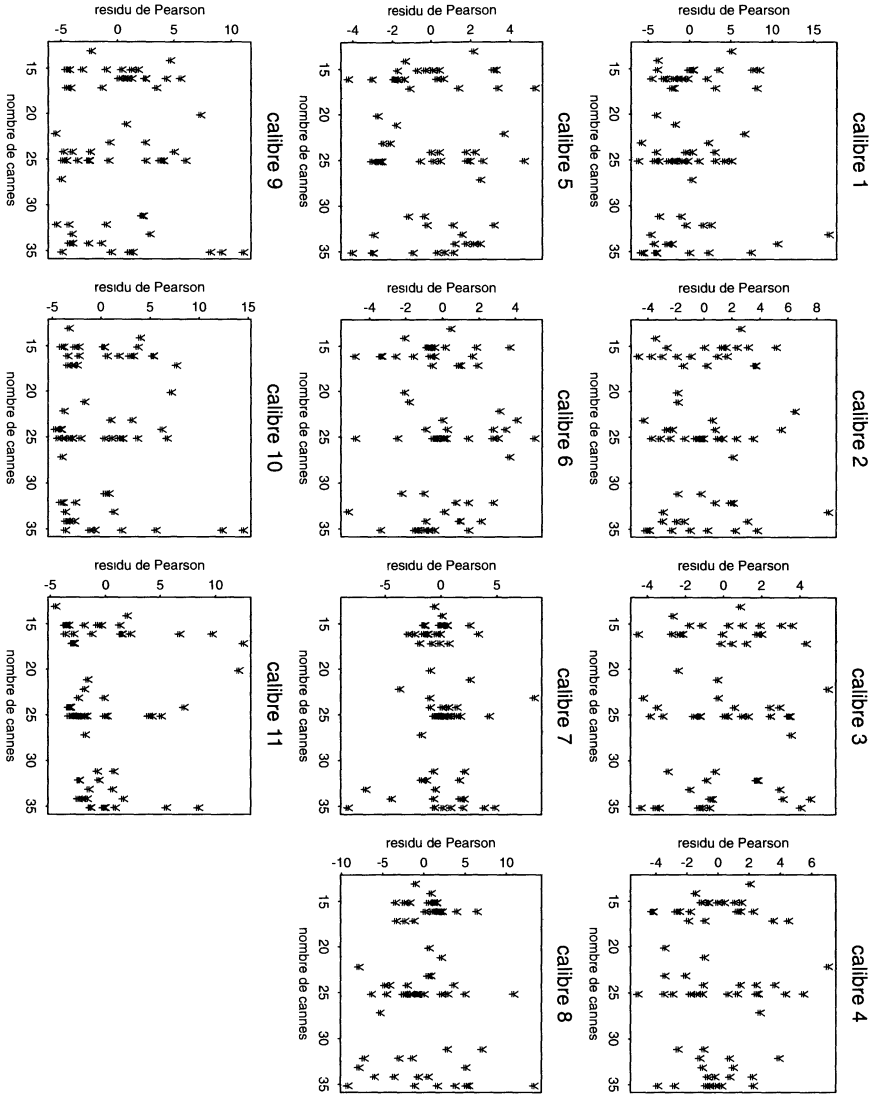

FIGURE 3

Graphique des résidus de Pearson en fonction du nombre de cannes pour chaque classe de calibre pour le modèle multinomial simple.

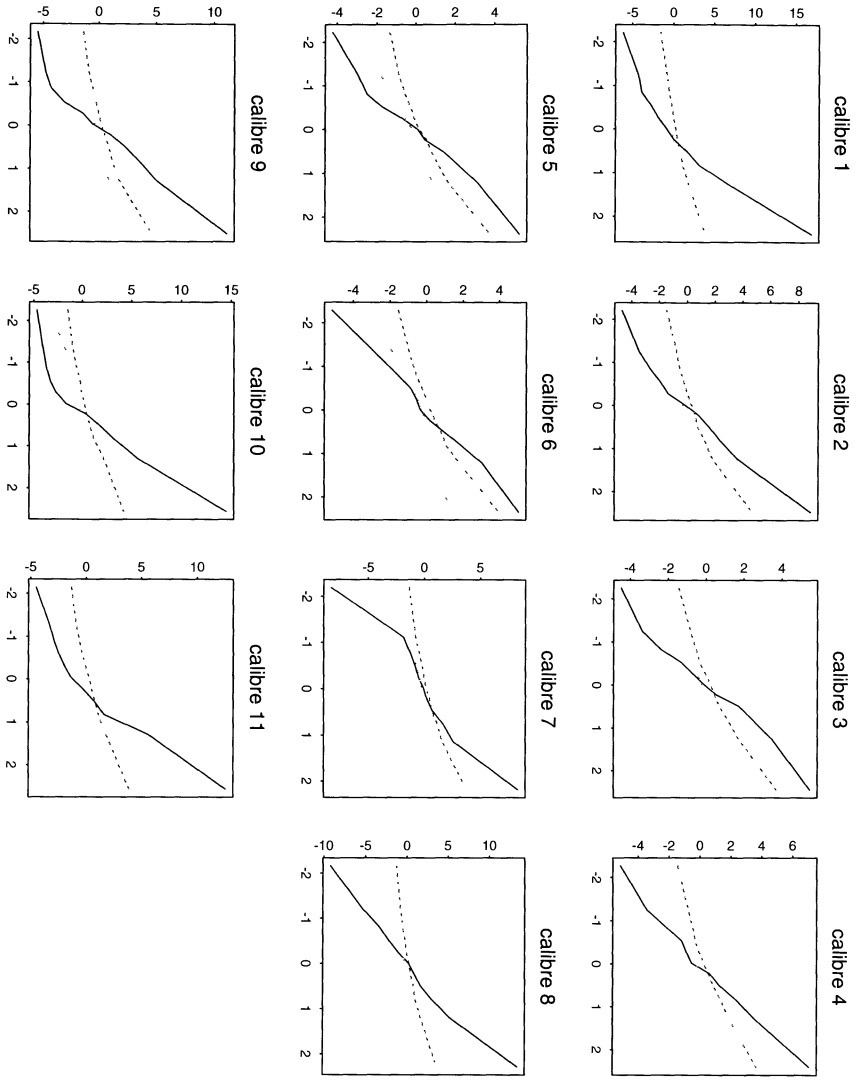

FIGURE 4

Représentation de type quantile-quantile pour les résidus de Pearson avec le modèle multinomial. Pour obtenir ces représentations on réalise des ajustements et on calcule des résidus sur 99 tableaux de données simulées sous le modèle estimé. La courbe en trait plein a comme abscisse la moyenne des quantiles, pour une probabilité donnée, des résidus obtenus par les simulations, l'ordonnée est le quantile correspondant pour les résidus de l'ajustement avec les données observées. Pour les courbes en traits pointillés l'ordonnée est le minimum (courbe en bas) ou le maximum (courbe en haut) des quantiles des résidus obtenus par les simulations.

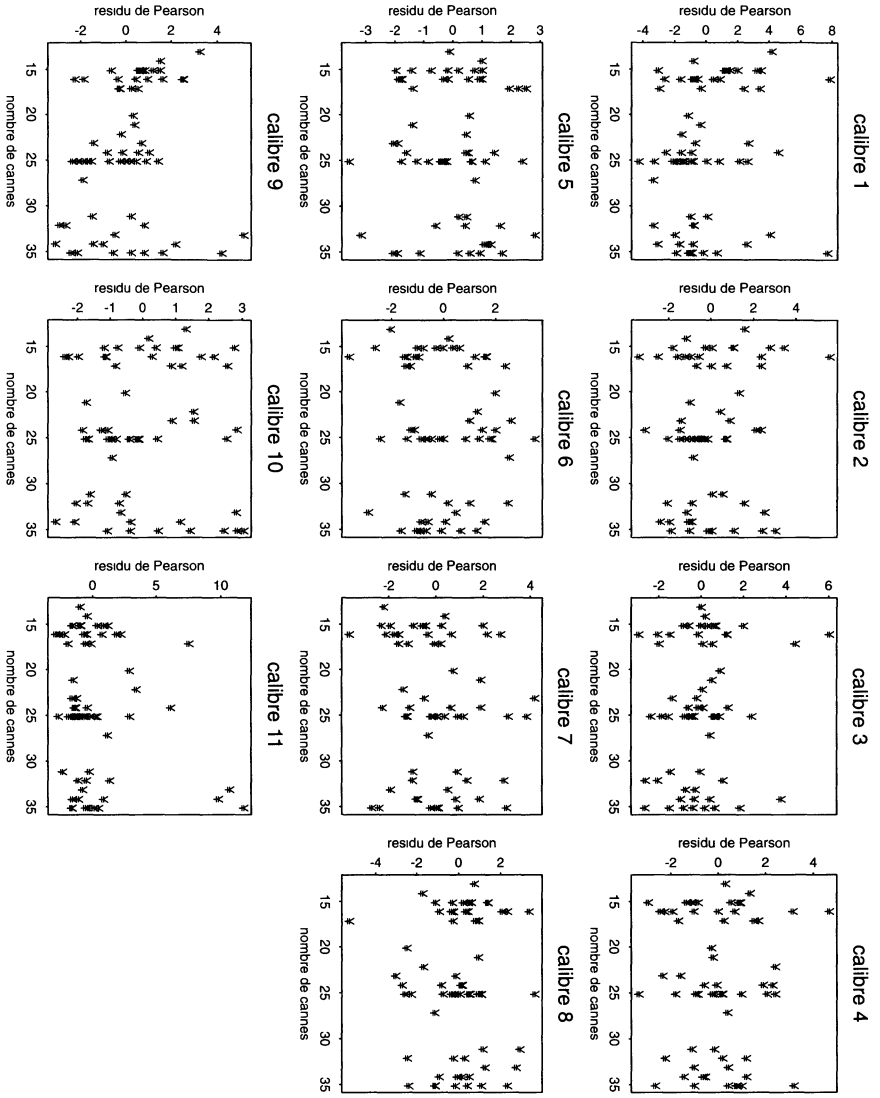

FIGURE 5

Graphique des résidus de Pearson en fonction du nombre de cannes pour chaque classe de calibre pour le modèle multinomial avec effet liane.

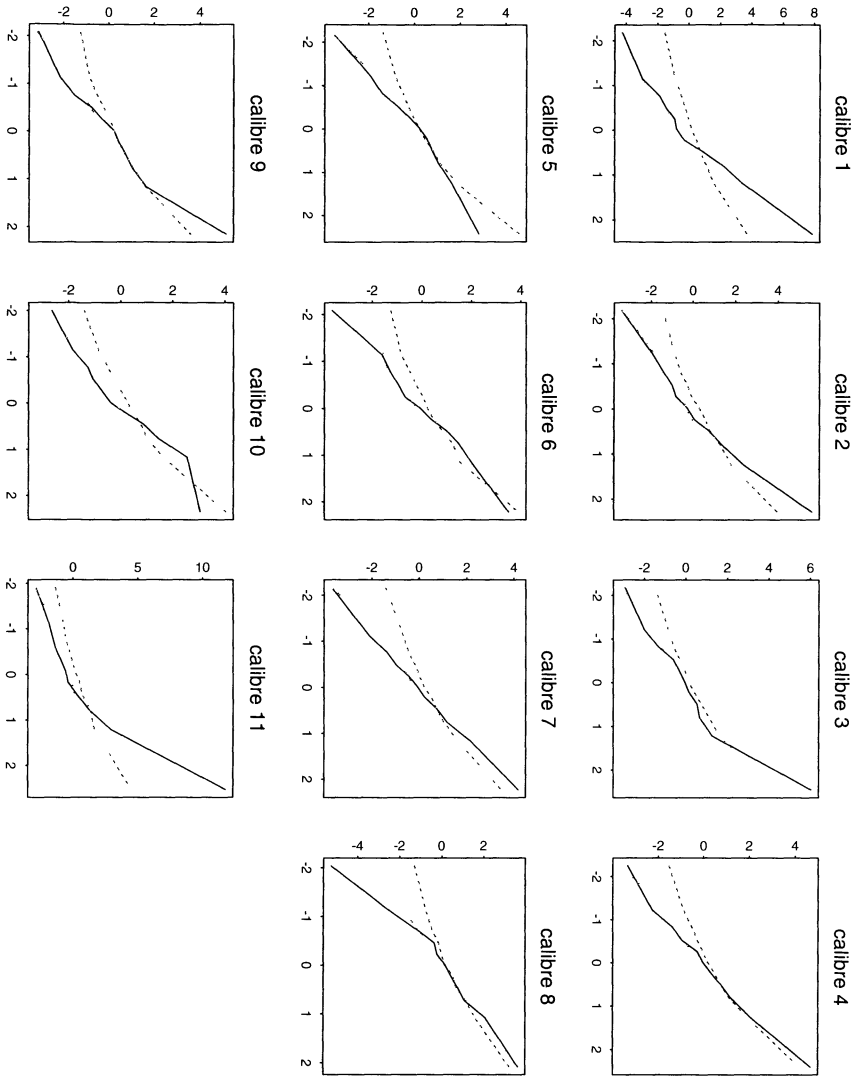

FIGURE 6

Représentation de type quantile-quantile pour les résidus de Pearson avec le modèle multinomial avec effet liane. Les courbes sont de même nature que sur la figure 4.

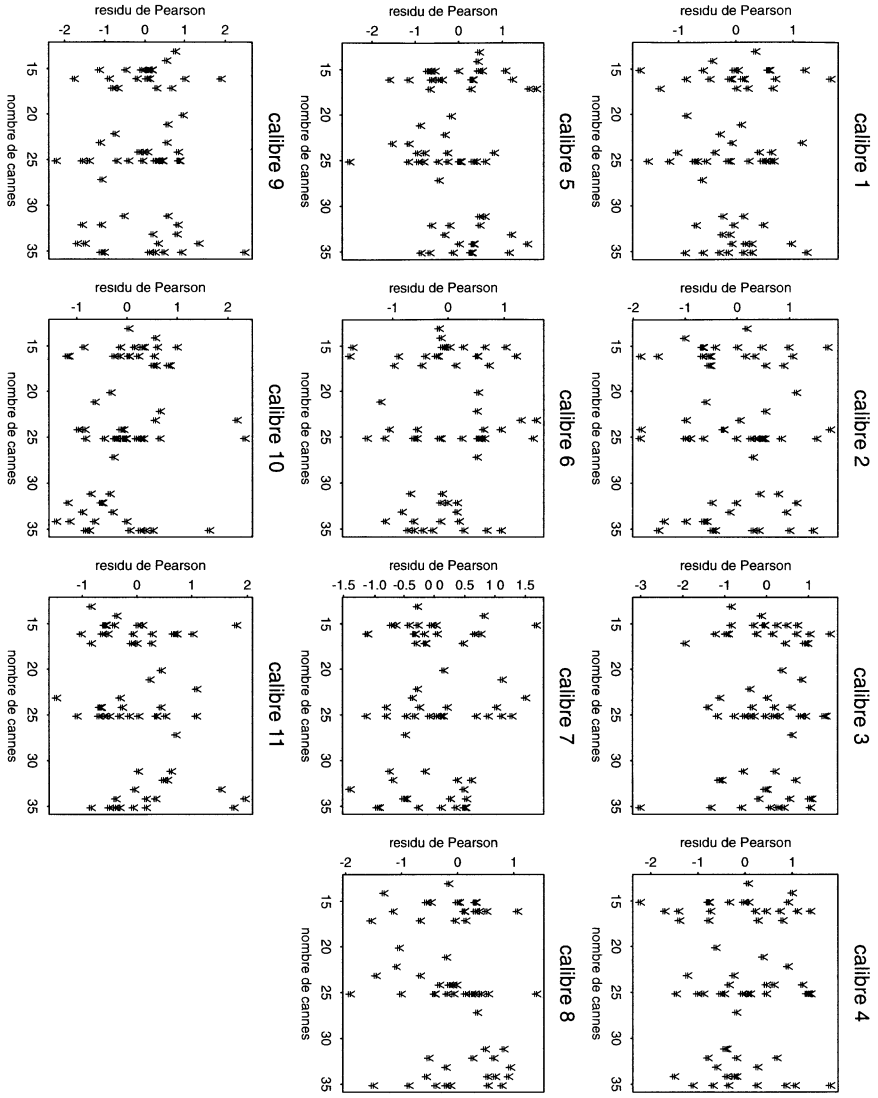

FIGURE 7

Graphique des résidus de Pearson en fonction du nombre de cannes pour chaque classe de calibre pour le modèle multinomial à effet aléatoire.

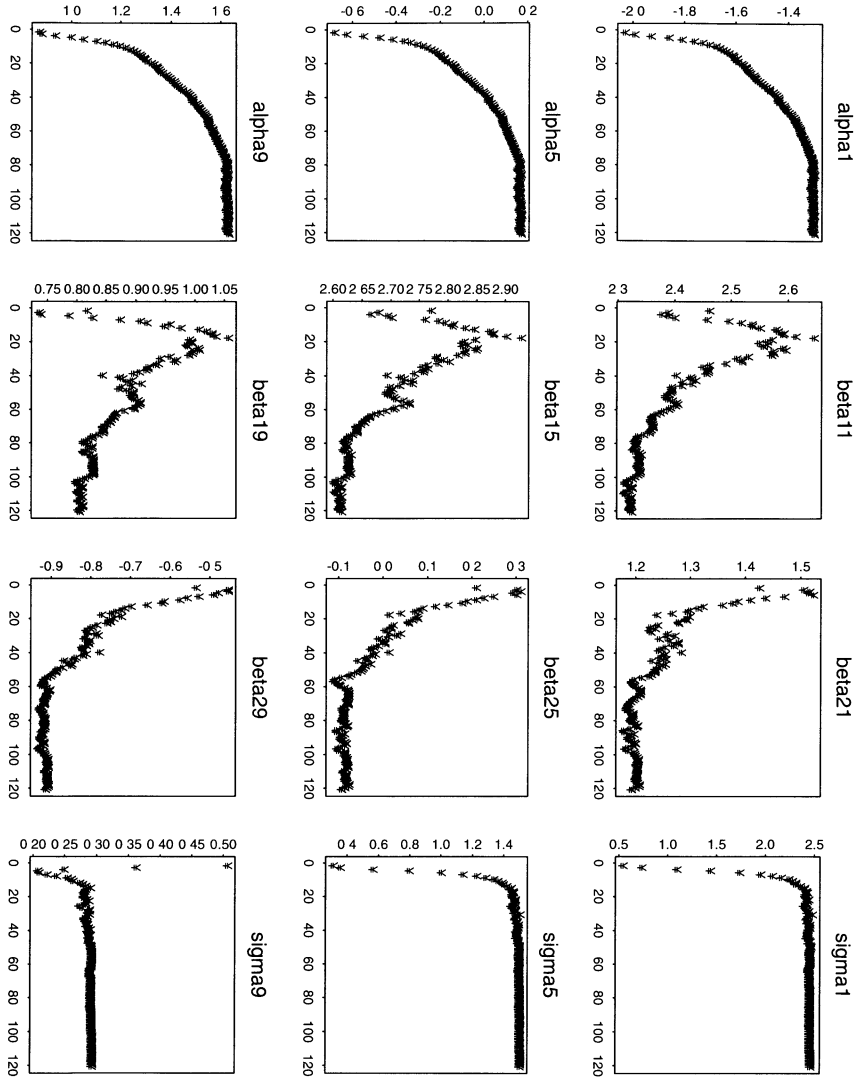

FIGURE 8

Graphique montrant l'évolution de certains paramètres du modèle à effet aléatoire ans le processus itératif.

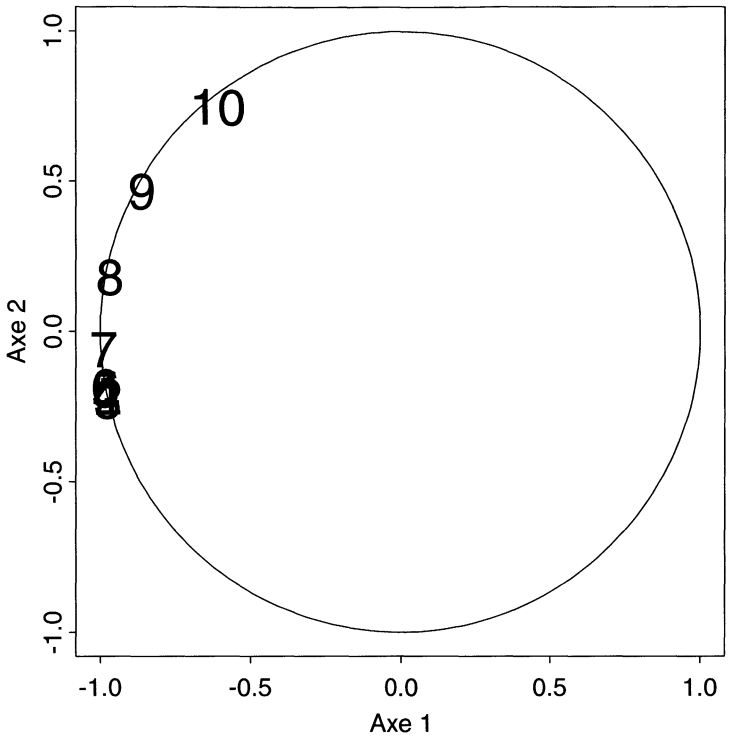

FIGURE 9

Représentation des composantes aléatoires dans le cercle des corrélations après une analyse en coordonnées principales à partir de la matrice de corrélations estimée.

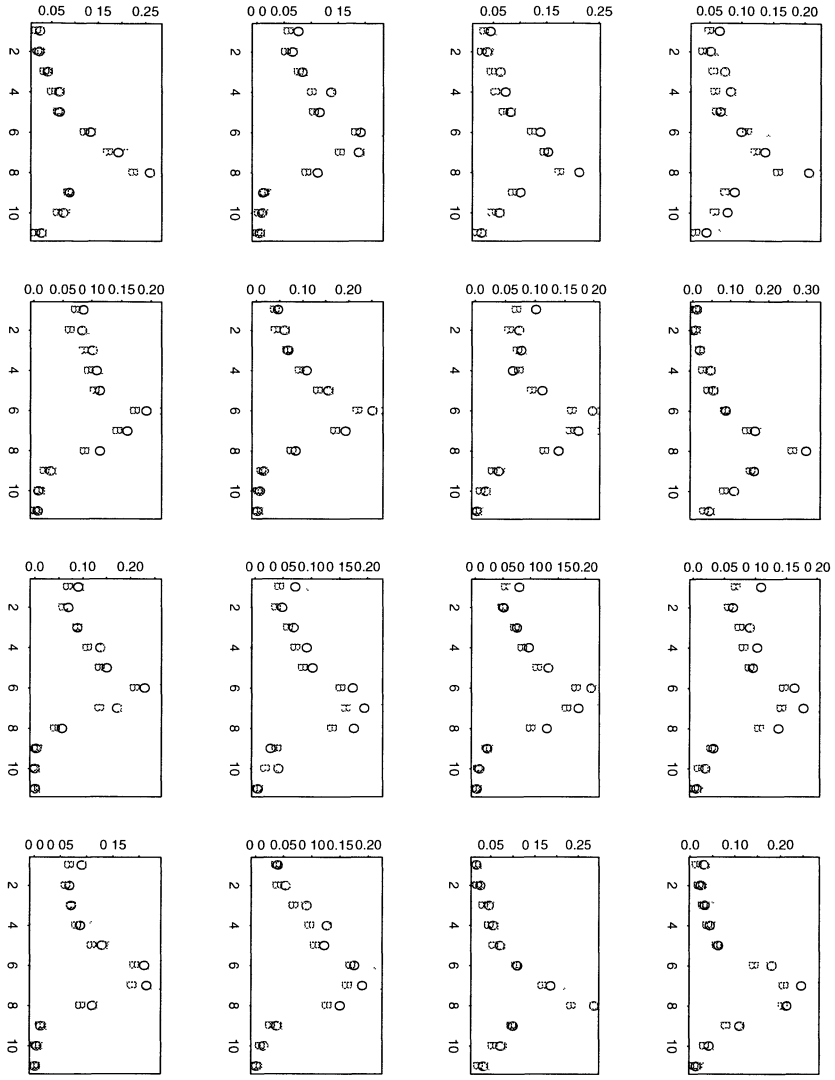

FIGURE 10

Estimation par intervalle des probabilités de chaque catégorie pour les 16 premières lianes en utilisant la loi a posteriori. *Pour chaque catégorie « 0 »est la fréquence observée, « H »(« B ») est la valeur maximale (minimale) de la probabilité parmi l'ensemble des valeurs obtenues en utilisant le modèle estimé et en générant 200 valeurs de l'effet aléatoire sous la loi* a posteriori.