Epidémiol. santé anim., 1997, 31-32

MODELISATION DE L'OCCURRENCE DES MAMMITES CLINIQUES AU SEIN D'UN TROUPEAU: ANALYSE DE LA SURDISPERSION

Gasqui P.[1], Coulon J.-B.[2]

For a study of clinical mastitis count in dairy cows within a lactation, using the poisson distribution is now classic. When the dependent variate is the interval between successive cases, it is also classic to use a parametric survival model with the exponential distribution. This two models assume independent observations. The violation of this assumption result in overdispersion within a herd.
The approach with survival model has permited to include in this model, a "biological dependance" parameter beetween successive cases within a lactation, in more individuals factors, time-dependent covariates and lactation length as censoring.
The data of INRA experimental herd have been analysed with this approach. The prediction of the number of cases per lactation or per animal was compared to observed number. The overdispersion seem explained essentially with this "biological dependance" parameter.

INTRODUCTION

Les mammites représentent le trouble sanitaire le plus fréquent en élevage laitier. La modélisation de leur occurrence constitue un outil précieux d'aide à la conduite des troupeaux. La construction d'un tel outil se heurte à un certain nombre de problèmes liés en particulier à la non-indépendance des mammites successives au cours de la lactation, à la non-indépendance des lactations successives au cours de la carrière d'une vache, et à la non-indépendance des vaches entre elles au sein d'un même troupeau. L'objectif de cette étude est de proposé une méthode pour résoudre ces perturbations. Pour illustrer cette approche nous avons choisi d'utiliser les données recueillies dans un troupeau expérimental de l'INRA, dans lequel les conditions de prévention sanitaire sont suffisamment bien respectées, pour rendre légitime l'hypothèse d'indépendance des vaches au sein du troupeau.

LE MODELE DE BASE

Chez la vache laitière en cours de production, le diagnostic de mammite clinique est réalisé au moment de la traite. Pour chacune de ces traites on observe la réalisation d'une variable aléatoire qui ne peut prendre que deux modalités : mamelle "atteinte" ou "indemne" d'une mammite clinique. Au cours d'une lactation la vache est soumise à un nombre important de traites successives. Si on fait l'hypothèse que ces traites sont indépendantes les unes des autres, et si on s'intéresse à la variable aléatoire "nombre de traite ayant donné lieu à un diagnostic de mammite clinique au cours de la production", on est naturellement amené à considérer que cette variable suit une distribution binomiale. Comme le nombre de traite est relativement important et que la probabilité de l'événement d'intérêt est relativement faible, on est dans le cas ou l'on peut approximer cette distribution binomiale par une distribution de poisson.

APPROCHE CLASSIQUE

Cette approche a consisté généralement à se placer dans le cadre de modèle GLM (Generalized Linear Model) (4); or avec ces modèles on est souvent confronté à la mise en évidence d'une "sur-dispersion", considérée par de nombreux auteurs (5) comme une mesure de l'écart aux hypothèses d'indépendance initiales. Les auteurs Diggle, Liang et Zeger (1) ont pour cela proposé la méthode GEE (Generalized Estimating Equations) pour estimer les corrélations entre les réponses au sein d'un groupe. On obtient malgré tout des paramètres assez difficilement interprétable en termes biologiques, et de toute façon une incapacité à rendre compte de la dépendance entre mammites successives au sein d'une lactation.

APPROCHE MODELE DE SURVIE

Quand le nombre d'événement indépendant survenant dans un certain intervalle de temps, suit une distribution de poisson, la variable aléatoire correspondant à l'écart de temps entre deux événements successifs suit une distribution exponentielle (2,3); l'instant d'occurence Ti du ième événement suit alors une distribution gamma de paramètre i.
Pour une vache donnée et pour une de ses lactations, si on a observé deux mammites aux instants t1 et t2, on dispose des écarts de temps : d1 (écart entre la date de vêlage t0 et t1), d2 (écart entre t1 et t2) et d* (écart entre t2 et la date de tarissement t*). Sous l'hypothèse d'indépendance des mammites successives, la vraissemblance s'écrit dans ce cas: V=f(d1).f(d2).S(d*), où f est la fonction de densité d'une loi exponentielle:

[1] Laboratoire d'Ecopathologie, I.N.R.A., 63122 Saint Genès Champanelle, France
[2] Laboratoire d'Adaptation des Herbivores aux Milieux, I.N.R.A., 63122 Saint Genès Champanelle, France

$f(d)=\lambda.\exp(-\lambda.d)$ pour tout $d\geq 0$, avec λ le nombre moyen d'événements par jour si les écarts d sont exprimés en jour; $1/\lambda$ est alors l'écart de temps moyen (en jour) entre 2 événements consécutifs, et S = 1-F, avec F pour la fonction de répartition d'une loi exponentielle. On a $S(d^*)=P(D\geq d^*)=\exp(-\lambda.d^*)$ pour tout $d^*\geq 0$, c'est-à-dire la probabilité de ne pas avoir observé de mammites durant l'intervalle d*; S est appelée "fonction de survie". Cela revient à considérer l'instant t* comme un instant de censure à droite. La vraissemblance prend donc en compte toute la durée de production, qu'il y ait eut ou non des mammites cliniques observées durant cette durée de production. Si on introduit la probabilité qu'une vache ait une mammite au temps d, conditionnellement au fait qu'elle n'en avait pas eu jusqu'à cette date d, on introduit la fonction de risque: $h(d)=f(d)/S(d)$ pour tout $d\geq 0$. Pour une distribution exponentielle cette fonction de risque vaut: $h(d)=\lambda$ pour tout $d\geq 0$, c'est-à-dire que le risque est constant au cours du temps et vaut λ.

Pour un ensemble de N* lactations avec N mammites, on a une vraissemblance qui s'écrit : $V=\prod_{i=1}^{N} f(d_i).\prod_{j=1}^{N^*} S(d_j^*)$, qui conduit à une estimation du risque: $\max_{\lambda}(V) \Rightarrow \frac{1}{\hat{\lambda}} = \frac{1}{N}.\sum_{j=1}^{N^*} T_j^*$, qui dépend des durées de production T* réellement observées pour les N* lactations.

Le modèle de base s'exprime donc simplement dans le cadre des modèles de Survie. On obtient un modèle qui dépend d'un paramètre λ qui a un sens en terme biologique, en ayant pris en compte les durées de production réellement observées.

Pour une lactation, si on suppose que le risque est constant au cours du temps t*, et si W est la variable aléatoire indiquant le nombre de mammite clinique observé durant ce temps t*, on retrouve le fait que W suit une loi de poisson de densité: $P(W=w)=\frac{(\lambda.t^*)^w}{w!}.\exp(-\lambda.t^*)$ avec $w\geq 0$, c'est-à-dire une loi de Poisson de paramètre $(\lambda.t^*)$. On parle alors de processus de Poisson homogène sur t*, c'est-à-dire à risque λ constant sur toute la période considérée.

Pour un ensemble de N* lactations ayant des durées de production différentes T*, et des caractéristiques qui font que le risque Λ associé à chaque lactation est différent, on aura alors affaire à un mélange de distributions de poisson, mélange qui généralement ne suit pas une distribution de poisson. Ce phénomène explique une partie de la sur-dispersion observée lors d'un ajustement d'un mélange par une distribution de Poisson (3). L'approche modèle de Survie permet de prendre en compte les durées effectivement observées, et les caractéristiques individuelles et d'élevage responsable de niveaux de risque différents.

Mais d'autres éléments biologiques font qu'au niveau du troupeau, on s'éloigne d'une distribution de poisson de base, et font apparaitre une sur-dispersion.

Existence de covariables dépendant du temps : processus de poisson homogène par intervalles.

Le risque peut aussi être considéré comme non constant au cours d'une lactation en raison de facteurs physiologiques (risque supérieur en tout début de lactation) ou environnementaux (risque supérieur à l'étable comparativement au pâturage). En considérant que le risque est constant au cours d'une période donnée on parle alors de processus de poisson homogène par intervalles. Les valeurs de ces covariables peuvent être constantes d'une lactation à l'autre (période post-vêlage et suite de la production) ou variables (par exemple selon la durée des périodes de stabulation et de pâturage).

Si par exemple, la durée de production est découpée en deux périodes successives de durée variable, et si on veut revenir à l'expression de la loi du nombre de mammite W sur la lactation, on fera intervenir les distributions de poisson que suivent respectivement W1, le nombre de mammite lors de la première période, et W2, le nombre de mammite lors de la deuxième période. Sur chacune de ces périodes le risque est considéré comme constant. On aura ainsi: $P(W=1) = P(W1=1).P(W2=0) + P(W1=0).P(W2=1)$; W ne suit pas une distribution de poisson.

Existence d'une dépendance entre mammites successives au cours d'une même lactation : mélange de distributions.

C'est le phénomène qui semble le plus important en terme de facteur de sur-dispersion. Le cadre des modèles de survie permet l'intégration dans le modèle d'un paramètre qui va prendre en compte une certaine dépendance entre ces mammites successives. On considère qu'à l'issu d'une mammite clinique, la vache va être soumise dans une proportion p des cas à un niveau de risque λ identique à celui auquel elle était soumise avant la mammite, alors que dans une proportion (1-p) des cas elle sera soumise à un niveau de risque λr (risque "récurrence") plus élevé que précédemment. Ce nouveau niveau de risque sera indépendant des covariables dépendant du temps et des facteurs différenciant les lactations. Cela revient par exemple à considérer que le traitement effectué lors d'une mammite clinique a simplement supprimé les symptômes de la mammite sans pour autant conduire à l'élimination complète des germes de la mamelle.

Le niveau de risque suivant une mammite est donc réévalué à chaque occurence d'une mammite clinique. Si W est le nombre de mammite durant la période de durée "u", et si la première mammite a lieu durant cette période, on a par exemple: $P(W=1|u) = p.\lambda.u.\exp(-\lambda.u) - \frac{(1-p)}{(\lambda.u-\lambda_r.u)}.\lambda.u.\left(\exp(-\lambda_r.u)-\exp(-\lambda.u)\right)$; W ne suit pas une distribution de poisson.

APPLICATION DE LA METHODE A UN TROUPEAU EXPERIMENTAL DE L'INRA

On a utilisé les données de 427 vaches, Française Frisonnes ou Montbéliardes ayant réalisé 1256 lactations dont la durée était comprise entre 180 et 500 jours. On dispose de toutes les mammites cliniques observées au cours de la production sur ces vaches, soient 344 mammites, durant une période allant de 1967 à 1988 ; la répartition des mammites par lactation est donnée au tableau 1.

Un ***modèle M1*** basé sur un processus de Poisson homogène dans le cadre d'un modèle GLM par exemple, conduit à des estimés moins dispersés que les observés, et à une estimation de la sur-dispersion qui vaut 1.21.

Un ***modèle M2*** prenant en compte des facteurs individuels: rang de lactation, race, potentiel de production et durées de production effectives, une covariable étable/pâturage et une covariable séparant la durée de production en une période post-vêlage de deux jours contre le reste de la lactation, améliore légèrement les résultats (tableau I).

Dans le ***modèle M3***, la prise en compte de la dépendance sous la forme introduite précédemment, en plus des durées de production effectives, de la covariable étable/pâturage et de la période post-vêlage de deux jours, conduit à des estimations très proches des valeurs observées.

Tableau I

nbre de mammite / lactation	0	1	2	3	4+
nbre de lactation observé	981	217	50	5	3
M1: nbre de lactation estimé	956	260	36	3	0
M2: nbre de lactation estimé	969	239	41	6	1
M3: nbre de lactation estimé	986	225	39	3	3

Avec ce dernier modèle M3, on obtient en terme de risque λ, les estimations suivantes :

étable		pâturage			
2 premiers jours	suite lactation	2 premiers jours	suite lactation	risque "récurrence"	taux "récurrence"
0.0110	0.0012	0.0055	0.0004	0.0124	0.05

On constate ainsi que le risque à l'étable est bien plus élevé qu'au pâturage, que l'on a un risque plus élevé dans les deux premiers jours de la lactation, que le risque "récurrence" est du même ordre de grandeur qu'en début de lactation à l'étable, et surtout que le taux (1-p) de mammites donnant lieu à l'augmentation ultérieure du risque est relativement faible, de l'ordre de 5 %. Avec ce modèle M3, on prédit une distribution du nombre de mammites par vache, c'est-à-dire par carrière, comparable à ce que l'on a observé (tableau II).

Tableau II

nbre de mammite / carrière	0	1	2	3	4+
nbre de vache observé	227	118	43	25	14
nbre de vache estimé	228	120	51	18	10

Cela revient à dire que pour cette élevage, l'hypothèse d'indépendance des vaches entre elles, voire des lactations successives pour une même vache, hypothèses que l'on a implicitement prise semble tout à fait légitime. Cela est certainement à relier avec le faible taux de "récurrence" de 5 % estimé pour cet élevage.

CONCLUSION

Pour d'autres élevages l'hypothèse d'indépendance entre lactations successives est beaucoup moins justifiée. Le travail se poursuit donc actuellement dans le sens d'une prise en compte d'une dépendance entre lactations successives au niveau de la carrière de l'animal. Ce passage est rendu possible dans le cadre des modèles de survie introduit précédemment. La possibilité de prendre en compte directement dans le modèle différents aspects de la biologie de l'animal, montre le grand intérêt des modèles de survie dans l'étude des pathologies pouvant survenir de multiple fois au cours d'une lactation, voire au cours de la carrière de l'animal. On a ainsi la possibilité de modéliser de façon plus "biologique" la dépendance entre ces événements successifs, et donc d'expliquer la sur-dispersion observée.

REMERCIEMENTS

A Odile Pons (INRA, Biométrie, Jouy-en-Josas, France) pour son support théorique en ce qui concerne la prise en compte d'une dépendance entre événements successifs dans le cadre des modèles de Survie.

BIBLIOGRAPHIE

1. Diggle P.J., Liang K. et Zeger S.L., 1994, Analysis of Longitudinal Data, Clarendon Press.
2. Johnson N.L., Kotz S., 1970, Continuous univariate distribution, t. 1 et 2, John Wiley.
3. Johnson N.L., Kotz S., Kemp A.W., 1992, Univariate Discrete Distributions, 2nd édition, John Wiley.
4. McCullagh P., Nelder J.A., 1989, Generalized Linear Models, 2nd édition, Chapman & Hall.
5. McDermott J.J., Schukken Y.H., Shoukri M.M., 1994, Study design and analytic methods for data collected from clusters of animals, Preventive Veterinary Medecine, 18, 175-191.