

EVALUATION DE LA STRUCTURE DE CORRÉLATION POUR DES TEMPS D'INFECTION DE PIS DE VACHE À L'AIDE D'EXTENSIONS DES MODÈLES DE FRAGILITÉ CORRÉLÉS

Estelle Kuhn ¹ & Klaartje Goethals ² & Charles El-Nouty ³ & Luc Duchateau ⁴

¹ *INRA, MaIAGE, Domaine de Vilvert 78352 Jouy-En-Josas Cedex FRANCE
estelle.kuhn@jouy.inra.fr*

² *Department of Comparative Physiology and Biometrics, Faculty of Veterinary
Medicine Ghent University, Salisburylaan 133, B-9820 Merelbeke, BELGIUM
Klara.Goethals@UGent.be*

³ *University Paris 13 Sorbonne Paris Cité, LAGA, 99, avenue Jean-Baptiste Clément,
94340 Villetaneuse FRANCE elnouty@math.univ-paris13.fr*

⁴ *Department of Comparative Physiology and Biometrics, Faculty of Veterinary
Medicine, Ghent University, Salisburylaan 133, B-9820 Merelbeke, BELGIUM
Luc.Duchateau@UGent.be*

Résumé. Comprendre les relations entre les temps d'infection des pis de vache est une information essentielle pour contrôler les épidémies de mastitis dans les élevages. Nous proposons des extensions des modèles de fragilité corrélés pour évaluer et comparer différentes structures de corrélation entre les quatre pis de vache. L'algorithme stochastique SAEM-MCMC permet d'estimer rapidement les paramètres par maximum de vraisemblance dans ce type de modèle de fragilité complexe. Nous mettons en évidence l'existence d'une importante corrélation entre les temps d'infection, la corrélation étant plus importante entre deux pis situés tous les deux soit à l'avant, soit à l'arrière, qu'entre deux pis situés l'un à l'avant et l'autre à l'arrière. Cela signifie que l'infection d'un pis de vache est un facteur de risque pour les autres pis, en particulier lorsque les pis sont situés dans la même région frontale ou arrière.

Mots-clés. modèle de fragilité corrélé, algorithme EM stochastique, temps d'infection, mastitis, maximum de vraisemblance

Abstract. The association between infection times of the udder quarters of a dairy cow is essential information for control of udder infections in a dairy cow herd. Extensions of the correlated frailty model are proposed to investigate and compare different correlation structures amongst the four udder quarter infection times of clustered within a cow. Such complex frailty models can be fitted with the SAEM-MCMC algorithm. It is demonstrated that substantial correlation exists between the udder quarter infection times, with the correlation within front and rear udder quarters being larger than between front and rear udder quarters. This signifies that an infected udder quarter is a risk factor

for the other udder quarters, especially when the udder quarter is in the same region, i.e., front or rear.

Keywords. correlated frailty model, stochastic EM algorithm, infection time, mastitis, maximum likelihood estimate

1 Problématique

Comprendre les mécanismes des infections des pis de vache constitue un enjeu majeur en épidémiologie animale. En particulier appréhender finement les relations qui existent entre les temps d'infection des pis de vache est une information essentielle pour contrôler les épidémies de mastitis dans les élevages. En particulier, des caractères individuels comme le statut primipare de la vache ont-ils une influence sur la propagation d'une infection? La propagation d'une infection est-elle influencée par la géométrie spatiale des pis de vache? Pour répondre à ces questions, nous disposons d'un jeu de données de temps d'infection de pis de vache censurés observés au niveau du pis sur un troupeau de 1196 vaches présentés par Goethals et al. (2009). Différentes covariables sont également renseignées, en particulier le statut primipare ou non de la vache et la localisation du pis infecté à l'avant ou à l'arrière de l'individu.

2 Modélisation et estimation des paramètres

Nous proposons d'analyser ces données à l'aide de modèles de fragilité qui prennent en compte d'une part les covariables liées au statut primipare de la vache et à la position du pis, d'autre part les effets de covariance entre les variables de fragilité. En particulier, nous présentons des extensions des modèles de fragilité corrélés avec différentes structures de corrélation entre les quatre pis de vache (cf. Kuhn et al. (2016)).

De façon plus précise, nous considérons une population de pis de vache répartis sur 1196 vaches, chaque vache ayant 4 pis. Pour $1 \leq i \leq 1196$ et $1 \leq j \leq 4$, les temps d'infection et les temps de censure pour le j -ième pis de la vache i seront notés T_{ij} et C_{ij} respectivement. Nous définissons les variables aléatoires $Y_{ij} = \min(T_{ij}, C_{ij})$ et $\Delta_{ij} = 1_{\{T_{ij} \leq C_{ij}\}}$, où 1_A désigne la fonction indicatrice de l'ensemble A . Les observations disponibles sont les couples $(Y_{ij}, \Delta_{ij})_{1 \leq i \leq 1196, 1 \leq j \leq 4}$. Dans la suite, on note $\mathbf{Y} = (Y_{ij})_{1 \leq i \leq 1196, 1 \leq j \leq 4}$ et $\mathbf{\Delta} = (\Delta_{ij})_{1 \leq i \leq 1196, 1 \leq j \leq 4}$.

Pour analyser ces données, nous considérons des modèles de fragilité (cf. Duchateau et al (2008), Wienke (2011)). On désigne le vecteur de fragilité de dimension 4 par $b_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4})^t$ pour la vache i pour $1 \leq i \leq 1196$ et le vecteur de l'ensemble des fragilités par $\mathbf{b} = (b_i)_{1 \leq i \leq 1196}$. Nous faisons les hypothèses classiques suivantes : les temps de censure $(C_{ij})_{1 \leq i \leq 1196, 1 \leq j \leq 4}$ sont indépendants des temps d'infection $(T_{ij})_{1 \leq i \leq 1196, 1 \leq j \leq 4}$

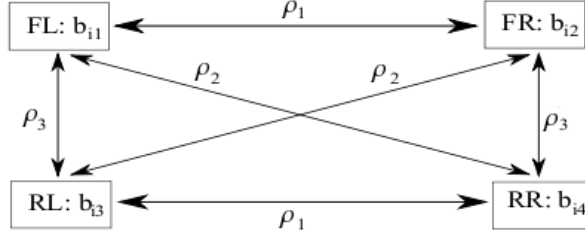


Figure 1: Représentation de la structure de corrélation pour les variables de fragilité des pis pour le modèle \mathcal{M}_1 où 'FL', 'FR', 'RL' et 'RR' désignent respectivement les pis avant gauche, avant droit, arrière gauche et arrière droit.

et des variables de fragilité $(b_i)_{1 \leq i \leq 1196}$; conditionnellement aux variables de fragilité $(b_i)_{1 \leq i \leq 1196}$, les temps d'infection $(T_{ij})_{1 \leq i \leq 1196, 1 \leq j \leq 4}$ sont indépendants.

On désigne par $\lambda_0(t)$ le risque de base au temps t . On désigne par x_{ij} le vecteur de covariables composé du statut primipare ou non de la vache et de la localisation avant ou arrière du pis. On désigne par $\beta = (\beta_1, \beta_2)$ le vecteur correspondant aux effets de ces deux covariables.

Soit $\lambda_{ij}(t|b_i)$ le risque instantané conditionnel pour le pis j de la vache i au temps t . Le modèle de fragilité \mathcal{M}_1 , pour $1 \leq i \leq 1196$, $1 \leq j \leq 4$ et $t \geq 0$ est donné par :

$$\lambda_{ij}(t|b_i) = \lambda_0(t) \exp(x_{ij}^t \beta + b_{ij}), \quad (1)$$

où les variables de fragilité $(b_i)_{1 \leq i \leq 1196}$ sont indépendantes et identiquement distribuées selon une loi gaussienne multivariée centrée de matrice de covariance Σ donnée par :

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_3 & \rho_2 \\ \rho_2 & \rho_3 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{pmatrix}$$

où $\sigma^2 \geq 0$ et $(\rho_1, \rho_2, \rho_3) \in [-1, 1]^3$.

Le choix de cette matrice de covariance est guidée par la géométrie naturelle des pis de vache. Une représentation de la structure de corrélation est donnée par la Figure 1.

Nous considérons par ailleurs deux types de risque de base, un risque paramétrique de type Weibull et un risque constant par morceaux pour s'affranchir de l'influence sur nos résultats de toute hypothèse paramétrique au niveau du risque de base. Nous analysons également plusieurs sous-modèles avec des structures de covariance plus particulières.

Estimer les paramètres d'un modèle de fragilité repose généralement sur la vraisemblance observée et sur l'estimation du maximum de vraisemblance associé. Mais le calcul

direct est souvent difficile, voir impossible dès que le modèle est complexe. Puisque les variables aléatoires de fragilité ne sont pas observées, on se situe dans le cadre général des modèles à variables latentes. Pour approcher l'estimateur du maximum de vraisemblance observée dans de tels modèles, on a généralement recours à l'algorithme Expectation Maximization (EM) proposé par Dempster et al. (1977). Cependant, il n'est pas applicable dans beaucoup de cas pratiques et doit souvent faire l'objet d'approximations. Comme nous considérons des modèles de fragilité particulièrement complexes du fait de la structure de corrélation entre les variables de fragilité vectorielles, nous appliquons l'algorithme EM stochastique proposé par Kuhn et El Nouty (2013). Nous réalisons des expériences sur des données simulées. Nous appliquons ensuite les méthodes proposées au jeu de données réelles de mastitis présenté par Goethals et al. (2009).

3 Analyses et résultats

Nous estimons les paramètres des modèles, ainsi que la matrice d'information de Fisher associée. Nous comparons des modèles avec différentes structures de corrélation à l'aide d'un test du rapport de vraisemblance. Nous évaluons la pertinence de l'approximation asymptotique du test du rapport de vraisemblance dans notre contexte par une étude de simulation. Nous analysons également l'influence du choix de modélisation du risque de base sur les données réelles de mastitis. Nous mettons finalement en évidence l'existence d'une corrélation entre les temps d'infection, la corrélation étant plus importante entre deux pis situés tous les deux soit à l'avant, soit à l'arrière, qu'entre deux pis situés l'un à l'avant et l'autre à l'arrière. Cela signifie que l'infection d'un pis de vache est un facteur de risque pour les autres pis, en particulier lorsque les pis sont situés dans la même région frontale ou arrière.

Bibliographie

- [1] Dempster, A. P., Laird, N. M. et Rubin, D. B. (1977), Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society Serie B*, 39, 1, 1–38.
- [2] Duchateau, L. et Janssen, P. (2008), *The Frailty Model*, Statistics for Biology and Health, Springer-Verlag, New York.
- [3] Goethals, K., Ampe, B., Berkvens, H., Laevens, H., Janssen, P. et Duchateau, L. (2009), Modeling Interval-Censored, Clustered Cow Udder Quarter Infection Times Through the Shared Gamma Frailty Model, *Journal of Agriculturam, Biological, and Environmental Statistics*, 14, 1, 1–14.
- [4] Kuhn, E. et El-Nouty, C. (2013), On a convergent stochastic estimation algorithm for frailty models, *Statistics and Computing*, 23, 3, 413–423.

- [5] Kuhn, E., Goethals, K., El-Nouty, C. et Duchateau, L. (2016), Assessing the correlation structure in cow udder quarter infection times through extensions of the correlated frailty model, in revision.
- [6] Wienke, Andreas (2011), *Frailty models in survival analysis*, Chapman & Hall/CRC Biostatistics Series, CRC Press, Boca Raton, FL.