

REVUE DE STATISTIQUE APPLIQUÉE

P. LHERMINIER

Solution de Bayes du problème des déterminations de paternité chez les bovins

Revue de statistique appliquée, tome 24, n° 4 (1976), p. 43-60

http://www.numdam.org/item?id=RSA_1976__24_4_43_0

© Société française de statistique, 1976, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

SOLUTION DE BAYES DU PROBLÈME DES DÉTERMINATIONS DE PATERNITÉ CHEZ LES BOVINS

P. LHERMINIER

Institut Technique d'Élevage Bovin

RESUME. —

La détermination de la paternité, chez les bovins, dans le cas de deux saillies par deux pères distincts, est toujours possible par l'analyse des groupes sanguins ; cependant cette méthode est coûteuse et une solution de Bayes est recherchée permettant soit de rendre minimum le taux d'erreurs d'attribution, soit, ce taux étant fixé, de rendre minimum les cas d'indécision.

Le taux d'erreur à ne pas dépasser est 1 %. On a déterminé les probabilités de fécondation a priori de chacun des taureaux, en fonction de l'écart entre saillies.

Les fonctions de décision qui minimisent les risques ont été établies et montrent qu'on doit distinguer deux cas.

1) L'intervalle entre saillies est ≥ 22 jours.

La décision de paternité est toujours possible, l'erreur moyenne reste inférieure à 1 %.

2) L'intervalle entre saillies est < 22 jours.

Une zone minimum d'indécision est définie telle que les cas à analyser soient le moins nombreux possible et que le taux d'erreur soit égal à 1 % dans les zones de décision.

Un graphique pratique destiné aux Organisations d'Élevage a été établi ; on constate que l'utilisation en routine de ce graphique permet de résoudre 90 % environ des cas de double paternité tout en maintenant un taux d'erreur voisin de 1 %.

INTRODUCTION

L'état-civil des animaux domestiques se caractérise par la nature essentiellement statistique de l'information qu'il contient. Les cas où la filiation exacte d'un individu est exigée sont rares, reproducteurs destinés à l'exportation et aux concours, taureaux de Centres d'Insémination Artificielle, ne représentant pas plus de un millième des naissances.

Pour tout le reste du cheptel on se préoccupe moins de savoir s'il y a ou non erreur, que de fixer un taux d'erreur tolérable et de ne pas le dépasser.

Les sondages effectués depuis 10 ans dans les fichiers d'état-civil bovin (LHERMINIER, 1973) par contrôle de filiations selon la méthode des groupes sanguins, montrent un taux moyen d'erreur de filiation de 5 % pour l'ensemble de la France. On observe qu'un taux d'erreur de 1 %, correspondant aux meilleurs fichiers français, paraît être le minimum en deçà duquel il n'est pas possible de descendre.

Mots-clés : Elevage. Décision. Paternité. Théorème de Bayes.

Les erreurs de père représentent à elles seules 90 % au moins des erreurs d'état-civil ; 10 % sont dues à la non reconnaissance des veaux chez les éleveurs ; les erreurs de mère sont très rares (LHERMINIER, 1973).

Il existe deux façons de contrôler la paternité d'un veau :

— S'assurer, connaissant la date de saillie, la date de naissance et la durée moyenne de gestation de la race, que les dates de naissance prévues et observées coïncident, à un intervalle de confiance près. C'est la méthode utilisée journellement par toutes les organisations professionnelles chargées de l'état-civil bovin. Toutefois elle ne permet que de confirmer une paternité supposée connue à l'avance par l'éleveur. Si par contre des doutes existent sur la date de saillie ou sur le taureau, cette méthode peut conduire à des erreurs de décision.

— Appliquer la méthode des groupes sanguins, qui est réalisée en France par le Laboratoire de Génétique Biochimique, C.N.R.Z., 78350 Jouyen Josas. Par la confrontation des groupes sanguins du veau et de ses parents, on effectue des exclusions de paternité avec une certitude totale (LEFEBVRE, 1967). Inversement on peut affirmer qu'une filiation est "compatible", c'est-à-dire qu'elle a une certaine vraisemblance d'être juste et, habituellement on la considère comme telle. On utilise chez les bovins 11 systèmes sanguins définis par des réactions immunologiques et pour chacun d'eux les allèles sont très nombreux (de l'ordre de 450 pour le système B).

On connaît également dans le sang un grand nombre de substances qui présentent un polymorphisme génétique aisé à déterminer et dont les lois de transmission héréditaire sont connues. La diversité des combinaisons rend très improbable l'identité sérologique et immunologique de deux individus. Pour cette raison, les filiations "compatibles", sans être "certaines" ont une vraisemblance très supérieure à celle obtenue par le contrôle des durées de gestation. La probabilité d'affirmer "compatible" une filiation fautive est de l'ordre de 10^{-3} à 10^{-6} contre 5 % d'erreurs en moyenne dans les fichiers des organisations professionnelles. Dans la suite de l'exposé nous considérons comme "quasi-certaines" les filiations reconnues "compatibles" par le laboratoire.

Sur un cheptel d'environ 10 000 000 de femelles adultes, il y a en France plus de 7 000 000 de vaches inséminées par an (I.A. premières). Pour 1/3 d'entre elles une seconde insémination est requise, et éventuellement d'autres, ce sont les "I.A. de retour en chaleur", qui représentent entre 3 et 4 000 000 par an. Dans tous ces cas si les deux saillies consécutives ont été effectuées par deux taureaux différents, un doute subsiste sur la véritable paternité, on dit qu'il y a deux pères supposés. Cette situation est très fréquente et constitue une des sources d'erreurs de filiation les plus importantes.

L'analyse des groupes sanguins serait évidemment la méthode de choix, cependant elle est trop laborieuse pour être généralisée. La date de naissance reste la méthode la plus commode pour décider lequel des deux est le père véritable ; mais cette méthode, qui conduit à un taux d'erreur de quelques pourcent lorsque le père est a priori connu, risque de devenir impraticable lorsqu'on hésite entre plusieurs pères. Dans le cas de l'insémination artificielle ou de la "monte en main", où les dates de saillies sont connues exactement, une solution élégante serait une analyse statistique suffisamment fine pour pouvoir

décider des paternités vraisemblables sans dépasser un seuil d'erreur raisonnable et en ne laissant dans l'incertitude (recours aux groupes sanguins) qu'un nombre de cas minimum.

RENDEL (1958) étudiant le problème des doubles saillies, estime les probabilités de chacun des taureaux en fonction de l'intervalle entre saillie et de la date du vêlage. Mais il se place dans l'hypothèse selon laquelle les probabilités *a priori* des saillies 1 et 2 d'être fécondantes, sont égales ; c'est-à-dire que les "coefficients de Bayes" sont arbitrairement fixés à 0,5 et 0,5 chacun.

La solution intuitive qu'il propose a servi de réglementation empirique dans la plupart des organisations d'élevage responsables d'un fichier d'état-civil.

Toutefois la définition exacte des limites de décision n'est pas connue, ni le taux d'erreur qu'entraîne l'utilisation de ces règles. Enfin on constate que la proportion des cas indécis, qui doivent être soumis aux groupes sanguins est élevée, ce qui est gênant et coûteux pour l'éleveur (RENDEL, BOUW and SCHMID, 1962).

La solution que nous proposons tient compte des probabilités *a priori* de chaque taureau, donc les coefficients de Bayes peuvent être très différents de 0,5.

On se propose de définir des règles de décision permettant de choisir entre deux pères supposés, connaissant uniquement les dates de saillies et la date de naissance du veau, et d'estimer pour chaque décision les taux d'erreur correspondants. En outre ces taux d'erreur doivent être minimums et les cas d'indécision les plus rares possibles.

Ces contraintes, taux d'erreur et taux d'indécision minimum, constituent une solution de Bayes au problème des deux pères supposés : "In the terminology of statistical decision-functions, a Bayes solution is a decision function which minimises the average risk relative to some probability distribution" (Kendall and Buckland, p. 22).

MATERIEL ET METHODES

La résolution probabiliste du problème des deux pères supposés nécessite l'étude préalable de plusieurs paramètres. Les seuls événements connus sont : la date A_1 de la première saillie par le taureau 1, la date A_2 de la seconde saillie par le taureau 2, la date de vêlage V (fig. 1) observé.

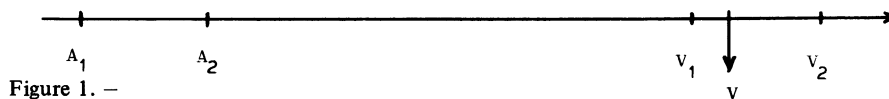


Figure 1. —

- A_1 : Dates de saillie par le taureau 1
- A_2 : Dates de saillie par le taureau 2
- V_1 : Dates de vêlage prévu correspondant à la saillie A_1
- V_2 : Dates de vêlage prévu correspondant à la saillie A_2
- V : Dates de vêlage observé

$$\Delta T = A_2 - A_1 = V_2 - V_1$$

$$G = V_1 - A_1 = V_2 - A_2 = \text{durée de gestation théorique.}$$

On doit donc étudier :

- La moyenne et la dispersion de la durée de gestation qui sépare la saillie fécondante du vêlage.
- Les probabilités a priori des taureaux 1 et 2 d'être chacun le véritable père, $P(A_1)$ et $P(A_2)$.
- Le seuil d'erreur à ne pas dépasser.

Durée de gestation.

Nous nous référerons à la revue présentée par BOUGLER et DERVEAUX (1969). Un certain nombre de facteurs interviennent pour modifier la durée de gestation, mais la plupart d'entre eux n'étant pas enregistrés dans la pratique de l'élevage, il n'a pas été possible de les retenir.

- Influence du milieu : effet de la région d'élevage
 effet de la saison de vêlage
 effet du mode d'exploitation
- Influence de l'âge de la vache
- Influence du format de la vache
- Influence du foetus : poids du veau
 sexe du veau
 veaux jumeaux
- Influences génétiques : les durées de gestation sont très variables selon les races et selon les types de croisement. Elles sont en général bien connues et font l'objet de mises au point périodiques par les Herd-Books de chaque race.

Le seul effet mesurable étant pratiquement celui de la race, nous accepterons pour durée de gestation moyenne, la valeur admise pour la race considérée, sans tenir compte des autres effets.

Loi de distribution des vêlages

Des difficultés importantes se rencontrent lorsque l'on veut tester la normalité de la loi des durées de gestation et estimer son écart-type. En effet la collecte de l'information est soumise à des biais de "choix" des dates de vêlage et de tri de durées de gestation considérées comme "anormales". La validation des Déclarations de Naissance (D.N.) est basée sur la vraisemblance de la durée de gestation, or les fourchettes d'acceptation ou de refus des D.N. ont été longtemps variables selon les organismes chargés de la collecte ; le calcul des variances montre une dispersion des valeurs qui reflète la sévérité ou l'indulgence des diverses réglementations.

Depuis 1969, il n'existe plus en France qu'une réglementation unique : fourchettes de + 2 et - 3 semaines par rapport au vêlage prévu ; bien entendu ceci ne résout pas le problème de l'écart-type réel et rend délicates les études de normalité. L'existence d'une dissymétrie dans la répartition des vêlages est d'autant plus gênante que nous aurons à comparer une branche droite avec une branche gauche.

D'autre part, les éleveurs connaissent la réglementation et il semble que certains en tiennent compte pour annoncer les dates de vêlage qui concordent avec les règles de validation des D.N.

RENDEL (1959) observe un effet de kurtosis significatif dans la répartition des durées de gestation ; toutefois il n'en tient pas compte dans les calculs ultérieurs. Nous accepterons nous aussi l'hypothèse de la normalité. L'écart-type est estimé à 5,6 par RENDEL. Une compilation portant sur 19 races françaises et 11 617 gestations recueillies par BOUGLER et DERVEAUX (1969) nous a permis d'obtenir l'estimation suivante de la variance :

$$\sigma^2 = 45,485 \quad \text{soit} \quad \sigma = 6,75 \text{ jours}$$

C'est cette valeur qui sera retenue dans les calculs. Pour des raisons de commodité on utilise plutôt la valeur $\sigma = 7$ jours = 1 semaine, dans la pratique.

Seuil d'erreur acceptable

Nous avons signalé en introduction qu'il était pratiquement impossible de déceler des taux d'erreurs de l'ordre de 1 % et que ces taux correspondaient aux meilleurs fichiers d'état-civil bovin français. Nous admettrons qu'une résolution du problème au seuil de 1 % sera satisfaisante.

Détermination des probabilités a priori de chaque père

Un premier échantillon de 572 cas de doubles paternités tranchées par groupes sanguins avait fait l'objet d'une étude préliminaire en 1972 et permis la construction d'une abaque adoptée par la Commission Nationale d'Amélioration Génétique et utilisée depuis par toutes les Organisations d'Elevage (LHERMINIER, 1972).

L'étude actuelle porte sur 884 cas de deux pères supposés et se propose de présenter l'analyse algébrique de la solution graphique déjà proposée ; le choix des cas analysés est malheureusement fortement biaisé puisque c'est en fonction d'un abaque provisoire que sont définis les litiges à résoudre.

Les 884 cas résolus ont été classés en fonction de l'intervalle ΔT entre deux saillies A_1 et A_2 par deux taureaux différents 1 et 2. On peut ainsi définir les probabilités a priori $P(A_1)$ et $P(A_2)$ que chacun des taureaux soit le véritable père, en fonction de l'écart entre les saillies.

Pour des raisons d'effectifs insuffisants et de commodité des calculs numériques nous avons regroupé certaines valeurs et accepté des arrondis (Tableau 1).

On note que parmi les cas proposés à l'analyse, la probabilité a priori du 1^{er} père est voisine de 1/2 lorsque quelques jours seulement le séparent du 2^e père, et qu'elle décroît progressivement pour atteindre 0,024 lorsque l'intervalle entre saillies est compris entre 17 et 26 jours ; ce qui correspond à la durée moyenne (21 jours) du cycle oestrien des bovins : la première saillie a été vaine, c'est la seconde qui est fécondante dans la grande majorité des cas. Beaucoup plus rarement la première saillie a été fécondante mais l'apparition de fausses chaleurs a conduit l'inséminateur à refaire une insémination dite "retour", inutilement.

Tableau 1

Estimation des probabilités a priori $P(A_1)$ et $P(A_2)$

Effectifs					
Intervalles ΔT	A_1	A_2	$P(A_1)$ observé	$P(A_1)$ retenu	$P(A_2)$
1	6	2	0,75	0,50	0,50
2	2	1	0,66		
3	1	1	0,50		
4	2	8	0,25		
5	2	0	0,50	0,25	0,75
6	1	6			
7	1	5	0,16		
8	2	8	0,25		
9	1	8	0,11		
10	1	6	0,14	0,10	0,90
11	2	12	0,14		
12	1	26	0,06		
13	2	5	0,28		
14	2	21	0,09	0,10	0,90
15	2	7	0,22		
16	1	22	0,04		
17	5	49	0,09		
18	1	91	0,01		
19	1	203	0,005	0,024	0,976
20	5	145	0,033		
21	1	88	0,011		
22	2	75	0,027		
23	1	27	0,037		
24	2	15	0,06		
25	0	12			
26	0	4			
Total	<u>47</u>	<u>837</u>			
	884				

Lorsque deux saillies consécutives A_1 et A_2 par deux taureaux 1 et 2 sont séparées par un intervalle ΔT , il existe des probabilités a priori $P(A_1)$ et $P(A_2)$ pour que chacun des taureaux soit le père du veau. Pour des raisons de commodité on a regroupé certaines données et accepté des arrondis.

RENDEL, et al. (1962), sur des échantillons assez différents des nôtres, trouvent les probabilités a priori suivantes (résultats regroupés) :

$$P(A_1) = 0,22 \text{ pour } \Delta T \text{ de } 1 \text{ à } 3 \text{ jours}$$

$$P(A_1) = 0,12 \text{ pour } \Delta T \text{ de } 4 \text{ à } 11 \text{ jours}$$

$$P(A_1) = 0,06 \text{ pour } \Delta T \text{ de } 18 \text{ à } 24 \text{ jours.}$$

Les valeurs de $P(A_1)$ et $P(A_2)$, c'est-à-dire les probabilités a priori de fécondation par les taureaux 1 et 2 sont, dans les calculs qui vont suivre, des "coefficients de Bayes". On voit que leurs valeurs sont très différentes de 0,50, qui serait la valeur admise si cette étude préliminaire sur un échantillon et avec l'aide des groupes sanguins, n'avait été réalisée.

RESULTATS

Enoncé du problème

Deux saillies sont effectuées sur la même femelle par les taureaux 1 et 2 aux temps A_1 et A_2 séparés par l'intervalle $\Delta T = A_2 - A_1$.

Les dates de vélages V_1 ou V_2 sont prévues d'après la durée de gestation moyenne de la race ; les cas où les taureaux seraient de races différentes sont assez exceptionnels pour être négligés.

L'écart entre les dates de vélages calculées est évidemment $V_2 - V_1 = \Delta T$.

Nous nous proposons de fixer les limites L de part et d'autres desquelles la paternité du vêlage V sera attribuée à l'un des taureaux 1 ou 2, avec (fig. 1)

- 1) un risque d'erreur minimum, ou bien,
- 2) un risque d'erreur inférieur ou égal à 0,01, et dans ce cas,
- 3) un nombre de cas d'indécision minimum.

Les cas indécis sont ceux pour lesquels le risque d'erreur du test dépasserait 0,01.

Nous avons donc une cause 1 de probabilité $P(A_1)$ dont la conséquence V , se produit autour de V_1 suivant la loi gaussienne d'écart-type σ ou (exclusif).

Une cause 2 de probabilité $P(A_2)$ dont la conséquence V , se produit autour de V_2 , suivant la même loi.

Les probabilités $P(A_1)$ et $P(A_2)$ a priori, ont été estimées (Tableau 1).

On a toujours $P(A_1) + P(A_2) = 1$.

Une erreur d'attribution de paternité se produit lorsque le vêlage V a lieu du côté de la limite L attribué à l'autre père ; par exemple (Fig. 2a) si $V > L$, alors que le père véritable est 1, le test désignera 2.

L'erreur α s'exprime donc :

$$\alpha = P(A_1) P(V > L/A_1) + P(A_2) P(V < L/A_2)$$

Nous allons dans une première démonstration calculer les valeurs de L qui rendent α minimum.

Puisque $P(A_1)$ et $P(A_2)$ varient en fonction de ΔT , c'est ΔT qui sera la variable.

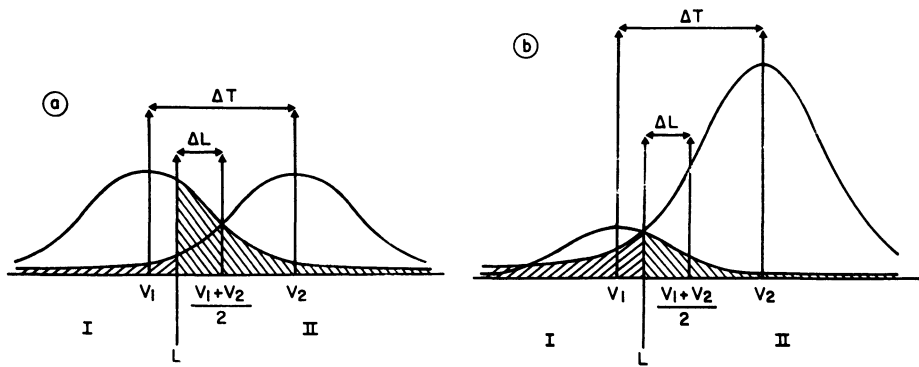


Figure 2 – Calcul de la limite L qui rend minimum l'erreur par inversion de pères (surface hachurée).

2a) La limite L ne se situe pas à mi-distance entre les dates de vélages prévues V_1 et V_2 . Elle s'écarte de $\frac{V_1 + V_2}{2}$ d'une durée ΔL estimée par la formule (2). Cet écart s'explique par le fait que les taureaux 1 et 2 n'ont pas des probabilités a priori égales, d'avoir fécondé la vache lors de la saillie.

2b) Si les répartitions des vélages autour de V_1 et V_2 ne sont plus a priori égales, mais pondérées par les probabilités a priori $P(A_1)$ et $P(A_2)$ de fécondation par l'un ou l'autre taureau, on remarque que la limite L se situe à l'intersection des deux courbes pondérées (relation 1).

Les événements situés à gauche de L (Zone I) sont rapportés au père 1, ceux situés à droite (Zone II) sont rapportés au père 2.

Résolution

1^{ère} partie

Recherche d'une limite L séparant les vélages avec une probabilité minimum d'erreur d'attribution.

La probabilité d'erreur est représentée graphiquement par la partie hachurée des figures 2a et 2b.

Les événements survenant à gauche de L sont retenus comme provenant de la cause 1, ceux situés à droite de la cause 2.

On peut écrire la loi de répartition des erreurs

$$\alpha = P(A_1) P(V > L/1) + P(A_2) P(V < L/2)$$

$$\alpha = P(A_1) \frac{1}{\sigma\sqrt{2\pi}} \int_L^{+\infty} e^{-\frac{(V-V_1)^2}{2\sigma^2}} dV + P(A_2) \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^L e^{-\frac{(V-V_2)^2}{2\sigma^2}} dV$$

On se propose de rendre α minimum

1/ On annule la dérivée

$$\frac{d\alpha}{dL} = P(A_2) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(L-V_2)^2}{2\sigma^2}} - P(A_1) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(L-V_1)^2}{2\sigma^2}} = 0,$$

(relation 1)

d'où :

$$e^{-\frac{(L - v_1)^2}{2\sigma^2} + \frac{(L - v_2)^2}{2\sigma^2}} = \frac{P(A_2)}{P(A_1)}$$

$$e^{\frac{v_2^2 - v_1^2 - 2L(v_2 - v_1)}{2\sigma^2}} = \frac{P(A_2)}{P(A_1)}$$

$$\frac{V_2 + V_1}{2} - L = \frac{\sigma^2}{V_2 - V_1} \log_n \frac{P(A_2)}{P(A_1)} \quad \text{avec} \quad \frac{V_2 + V_1}{2} - L = \Delta L$$

et
$$V_2 - V_1 = \Delta T$$

$$\boxed{\Delta L = \frac{\sigma^2}{\Delta T} \log_n \frac{P(A_2)}{P(A_1)}} \quad (\text{relation 2})$$

2/ On montre ensuite que cette valeur correspond au minimum de α c'est-à-dire que

$$\frac{d\alpha}{dL} > 0 \quad \text{pour} \quad \Delta L < \frac{\sigma^2}{\Delta T} \log_n \frac{P(A_2)}{P(A_1)}$$

et
$$\frac{d\alpha}{dL} < 0 \quad \text{pour} \quad \Delta L > \frac{\sigma^2}{\Delta T} \log_n \frac{P(A_2)}{P(A_1)}$$

$$\frac{d\alpha}{dL} = P(A_2) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(L - L_2)^2}{2\sigma^2}} \left[1 - \frac{P(A_1) e^{-\frac{(L - v_1)^2}{2\sigma^2}}}{P(A_2) e^{-\frac{(L - v_1)^2}{2\sigma^2}}} \right]$$

Le 1^{er} membre est toujours positif, le signe de la dérivée est toujours celui de la parenthèse.

1^{er} cas

$$\Delta L < \frac{\sigma^2}{\Delta T} \log_n \frac{P(A_2)}{P(A_1)}$$

d'après le calcul précédent

$$e^{-\frac{(L - v_1)^2}{2\sigma^2} + \frac{(L - v_2)^2}{2\sigma^2}} < \frac{P(A_2)}{P(A_1)}$$

donc
$$\frac{d\alpha}{dL} > 0$$

2^e cas

$$\Delta L > \frac{\sigma^2}{\Delta T} \log_n \frac{P(A_2)}{P(A_1)}$$

$$e^{-\frac{(L-v_1)^2}{2\sigma^2} + \frac{(L-v_2)^2}{2\sigma^2}} > \frac{P(A_2)}{P(A_1)}$$

La valeur trouvée pour ΔL est donc celle qui rend α minimum.

Remarque : D'un point de vue graphique si on construit non pas les 2 courbes gaussiennes identiques mais les courbes de probabilité pondérées, L se situe à l'intersection des deux courbes d'après la relation (1) (fig. 2b).

On remarque que tout déplacement de L de part et d'autre de l'intersection conduit à une augmentation de la surface hachurée totale donc de l'erreur α .

Il en résulte que le père choisi est toujours le plus vraisemblable.

3/ Calcul de la probabilité d'erreur α

on pose

$$F = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx$$

avec

$$x_1 = \frac{\frac{\Delta T}{2} - \Delta L}{\sigma} \quad \text{et} \quad x_2 = \frac{\frac{\Delta T}{2} + \Delta L}{\sigma}$$

$$\alpha = P(A_1) \left[1 - F\left(\frac{\frac{\Delta T}{2} - \Delta L}{\sigma}\right) \right] + P(A_2) F\left(\frac{\frac{\Delta T}{2} + \Delta L}{\sigma}\right) \quad \text{relation (3)}$$

2^{ème} partie

Cas où l'on fixe deux limites L_1 et L_2 (fig. 3).

Lorsque $\alpha > 0,01$ on définit une zone d'indécision III située entre les limites L_1 et L_2 .

On appelle : β la probabilité d'inversion de paternité, c'est-à-dire qu'un veau issu d'un père 1 soit né dans la zone attribuée au père 2 et inversement (zone hachurée de la figure 3).

γ la probabilité que l'on réserve sa décision (zone d'indécision, pointillée sur la figure 3).

La probabilité de décision est donc $1 - \gamma$

α' la probabilité d'erreur par inversion de paternité, ramenée aux cas décidés

$$\text{On a donc} \quad \alpha' = \frac{\beta}{1 - \gamma}$$

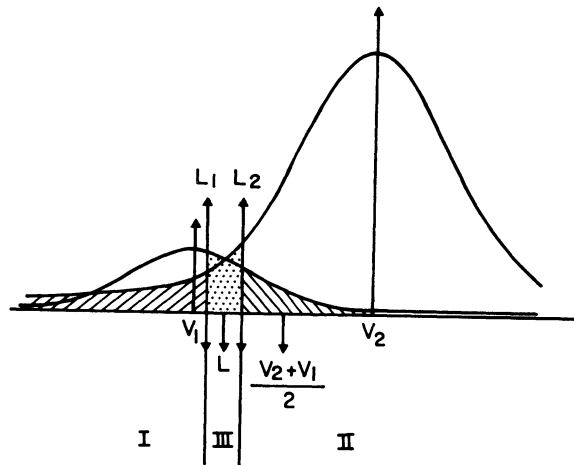


Figure 3 – Choix d'une zone d'indécision dans la cas où les erreurs par inversion sont trop fréquentes : lorsque la limite L entraîne une erreur $\alpha > 0,01$, on définit une zone d'indécision (Zone III, pointillée sur le graphique) limitée par L_1 et L_2 . Les zones I et II correspondent toujours aux taureaux 1 et 2, l'aire hachurée représente la probabilité d'erreur par inversion $\alpha' = 0,01$; l'aire pointillée correspond à des vélages dont on ne pourra déterminer la paternité que par l'analyse des groupes sanguins, cette surface doit être minimum pour avoir le moins de cas possible à soumettre à l'analyse.

On impose les contraintes suivantes :

Première contrainte : La probabilité d'erreur par inversion de paternité est constante par rapport aux cas décidés $\alpha' = 0,01$

Deuxième contrainte : La probabilité d'indécision doit être minimum de façon à ne soumettre à l'analyse des groupes sanguins que le moins de cas possible : γ minimum.

La première contrainte est satisfaite pour les couples (L_1, L_2) tels que

$$\frac{\partial B}{\partial L_1} dL_1 + \frac{\partial B}{\partial L_2} dL_2 + \alpha' \frac{\partial \gamma}{\partial L_1} dL_1 + \alpha' \frac{\partial \gamma}{\partial L_2} dL_2 = 0$$

soit

$$\frac{dL_1}{dL_2} = - \frac{\frac{\partial \beta}{\partial L_2} + \frac{\alpha' \partial \gamma}{\partial L_2}}{\frac{\partial \beta}{\partial L_1} + \frac{\alpha' \partial \gamma}{\partial L_1}}$$

ce qui définit un ensemble de points (L_1, L_2)

La seconde contrainte, γ minimum, implique

$$\frac{d\gamma}{dL_2} = \frac{\partial \gamma}{\partial L_1} \cdot \frac{dL_1}{dL_2} + \frac{\partial \gamma}{\partial L_2} = 0$$

$\frac{dL_1}{dL_2}$ étant défini par la première contrainte on obtient l'équation

$$\frac{-\frac{\partial \beta}{\partial L_2} \cdot \frac{\partial \gamma}{\partial L_1} + \frac{\partial \beta}{\partial L_1} \cdot \frac{\partial \gamma}{\partial L_2}}{\frac{\partial \beta}{\partial L_1} + \alpha' \frac{\partial \beta}{\partial L_1}} = 0$$

En remplaçant les dérivées partielles par leurs expressions on obtient :

$$\frac{d\gamma}{dL_2} = \frac{-P(A_1) e^{-\frac{(L_2 - V_1)^2}{2\sigma^2}} \cdot P(A_1) e^{-\frac{(L_1 - V_1)^2}{2\sigma^2}} + P(A_2) e^{-\frac{(L_2 - V_2)^2}{2\sigma^2}} \cdot P(A_2) e^{-\frac{(L_1 - V_2)^2}{2\sigma^2}}}{P(A_2) e^{-\frac{(L_2 - V_2)^2}{2\sigma^2}} + \alpha' P(A_1) e^{-\frac{(L_1 - V_1)^2}{2\sigma^2}}}$$

Cette fraction est toujours du signe du numérateur. Lorsque celui-ci est nul, on obtient la relation (4)

$$\boxed{\frac{V_2 + V_1}{2} - \frac{L_2 + L_1}{2} = \frac{\sigma^2}{\Delta T} \log_n \frac{P(A_2)}{P(A_1)}}$$

En comparant à la relation (2) on voit que $\frac{L_1 + L_2}{2} = L$ donc L_1 et L_2 sont symétriques par rapport à L .

Pour étudier le signe de $\frac{d\gamma}{dL_2}$ on met le numérateur sous la forme

$$\frac{d\gamma}{dL_2} = [P(A_2)]^2 e^{-\frac{(L_2 - V_2)^2 + (L_1 - V_2)^2}{2\sigma^2}} \left[1 - \frac{[P(A_1)]^2 e^{-\frac{(L_2 - V_1)^2 + (L_1 - V_1)^2}{2\sigma^2}}}{[P(A_2)]^2 e^{-\frac{(L_2 - V_2)^2 + (L_1 - V_2)^2}{2\sigma^2}}} \right]$$

Le signe de la dérivée est celui de la parenthèse. On établit immédiatement que la relation (4) rend γ minimum.

Il est remarquable que l'équation qui fixe une probabilité constante d'erreur α' soit aussi celle qui rende minimum le nombre de cas d'indécision.

Calcul des limites L_1 et L_2

On développe l'égalité $\beta = \alpha' (1 - \gamma)$

$$\text{soit } P(A_1) \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{L_1} e^{-\frac{(V - V_2)^2}{2\sigma^2}} dV + P(A_2) \frac{1}{\sigma \sqrt{2\pi}} \int_{L_2}^{+\infty} e^{-\frac{(V - V_1)^2}{2\sigma^2}} dV$$

On pose

$$F = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx$$

avec :

$$\Delta L_1 = L - L_1, \quad \Delta L_2 = L_2 - L \quad \text{et} \quad \Delta L_1 = \Delta L_2$$

$$x_1 = \frac{V_2 - L_1}{\sigma} = \frac{\frac{\Delta T}{2} + \Delta L + \Delta L_1}{\sigma}$$

$$x_2 = \frac{L_2 - V_1}{\sigma} = \frac{\frac{\Delta T}{2} - \Delta L + \Delta L_2}{\sigma}$$

$$x_3 = \frac{L_1 - V_1}{\sigma} = \frac{\frac{\Delta T}{2} - \Delta L - \Delta L_1}{\sigma}$$

$$x_4 = \frac{V_2 - L_2}{\sigma} = \frac{\frac{\Delta T}{2} + \Delta L - \Delta L_2}{\sigma}$$

On obtient alors la relation (5)

$$\beta = P(A_1) F(x_2) + P(A_2) [1 - F(x_1)] = \alpha'(1 - \gamma)$$

avec

$$\gamma = P(A_1) F(x_2) - P(A_1) F(x_3) + P(A_2) [1 - F(x_4) + F(x_1)]$$

qui permet d'estimer les valeurs numériques de β et de γ ainsi que les limites L_1 et L_2 .

DISCUSSION

Calculs numériques, présentation graphique des résultats

La détermination de la limite ΔL qui minimise l'erreur α est simplement donnée par la relation (3). Pour les différents intervalles ΔT entre vélages prévus on a calculé les valeurs correspondantes de ΔL .

Cependant, les valeurs numériques des probabilités a priori $P(A_1)$ et $P(A_2)$ ne varient pas de façon continue, donc les valeurs de ΔL non plus. Cet inconvénient est peu grave puisque le calcul de l'erreur minimum α montre que $\alpha > 0,01$ pour $\Delta T < 22$ jours ; les valeurs "utiles" de ΔL seront donc uniquement celles où $P(A_1)$ et $P(A_2)$ ont des valeurs constantes, lorsque $\Delta T \geq 22$ jours.

Lorsque $\Delta T < 22$ jours, $\alpha > 0,01$, on doit alors définir une zone d'incertitude avec les limites L_1 et L_2 . Suivant la relation (5) la détermination de ΔL et ΔL_1 se fait à l'aide d'une table de la loi normale.

Les limites doivent, pour des raisons pratiques, comporter un nombre entier de jours.

Il n'est donc pas nécessaire de rechercher une trop grande précision dans les calculs numériques. D'ailleurs pour cette partie du graphique (ΔT compris entre 10 et 17 jours) les fréquences $P(A_1)$ et $P(A_2)$ sont estimées sur des effectifs assez faibles et entachées d'une large incertitude (tableau 1). Les valeurs adoptées résultent de regroupements commodes : 0,1 – 0,25 – 0,50 qui facilitent les calculs assez laborieux des valeurs limites. Par contre, dans l'établissement du graphique d'utilisation pratique (fig. 4b) il n'y a aucune raison d'admettre de telles discontinuités et nonobstant les marches d'escalier qui figurent des jours entiers, nous avons dessiné une ligne régulière tout en conservant au mieux les surfaces des trois zones c'est-à-dire les probabilités respectives des deux pères ou de l'indécision.

D'autre part, une réglementation administrative invalide les vélages situés hors des limites de + 15 jours et – 21 jours de part et d'autre de la date prévue, il est donc inutile de rechercher des limites significatives supérieures à ces durées puisqu'elles tomberaient sous le coup de cette réglementation. Donc pour $\Delta T \leq 10$ jours aucune paternité ne peut être attribuée.

En vertu du même règlement pour $\Delta T > 36$ jours les zones permises pour les 2 pères ne se chevauchent plus, ils sont examinés indépendamment l'un de l'autre.

Seules les valeurs ayant une signification pratique ont été calculées (Tableau 2).

Pour α , les valeurs supérieures à 0,01 n'ont pas d'intérêt puisque c'est le seuil d'erreur que l'on ne veut pas dépasser.

Pour ΔL_1 , les valeurs supérieures à 15 jours n'ont pas été calculées puisqu'elles fixent des limites qui dépassent la réglementation administrative en vigueur.

Pour γ la proportion de filiations soumises à l'analyse des groupes sanguins dépasse largement 0,50 à partir d'un intervalle $\Delta T \leq 10$ jours. Pour des raisons de commodité pratique déjà évoquées, on a fixé à 10 jours la limite en deçà de laquelle l'analyse sanguine est toujours exigée. Ces cas sont d'ailleurs relativement peu nombreux : 7 % des cas de deux pères supposés.

Utilisation pratique du graphique (fig. 4a et 4b).

Un graphique permet de définir des zones I, II et III correspondant aux pères 1 et 2 ou aux cas d'indécision (fig. 4a). Il est remarquable de constater la correspondance avec le graphique 4b établi de façon empirique en 1972 et qui a fait l'objet de la circulaire ministérielle DPMEE/SE/4 - 104. Il est utilisé depuis cette époque dans tous les organismes professionnels responsables de l'Etat-Civil Bovin. Sur un échantillon de 572 cas, l'utilisation du graphique a permis de ne réaliser que 43 analyses au lieu de 572, tout en ne révélant que 7 erreurs dues à des gestations exceptionnelles. Le graphique permet donc une économie considérable des cas à analyser tout en maintenant l'erreur au taux acceptable pour les fichiers d'Etat-Civil Bovin.

Tableau 2

Calcul des limites permettant le choix entre les deux pères supposés,
et taux d'incertitude

ΔT (jours)	$P(A_2)$	ΔL (jours)	α	ΔL_1 (jours)	γ
30	0,976	5,6			
26	—	6,46	0,006		
25	—				
24	—	7	0,008		
23	—				
22	—	7,6	0,010	0	0
21	—				
20	—	8,4	0,013	1 à 2	0,008
19	—				
18	—	9,3	0,016	3	0,030
17	—				
16	0,90	6,25	0,06	8	0,22
15	—	6,66			
14	—	7,14		10	0,33
13	—				
12	—	8,33		14	
11	—				
10	—	10	0,09	16	0,54
9	—	9			
8	0,75	6			
7	—				
6	—				
5	—	10			
4	0,50	5	0,415		
3	—				
2	—	10			
1	—	20			
0	—	∞	0,5		

Légende du tableau 2

ΔT Intervalle entre saillies : A_1 et A_2 , donc entre les vélages prévus V_1 et V_2 .

$P(A_2)$ Probabilité a priori de fécondation par le second père, donc de vélage autour de V_2 .

ΔL Limite de séparation entre les zones attribuées à l'un et à l'autre père, minimisant les erreurs par inversion.

α Taux d'erreurs d'attribution par application de la limite L .

$\Delta L_1, \Delta L_2$ Limites de séparation entre les zones attribuées aux pères et la zone d'indécision la plus réduite possible (valeurs arrondies).

γ Proportion de cas indécis par rapport aux cas résolus, ou surface relative de la zone d'indécision.

Les valeurs de L, L_1 , et γ sont calculées à partir des relations 3 et 5, à l'aide d'une table de la loi normale.

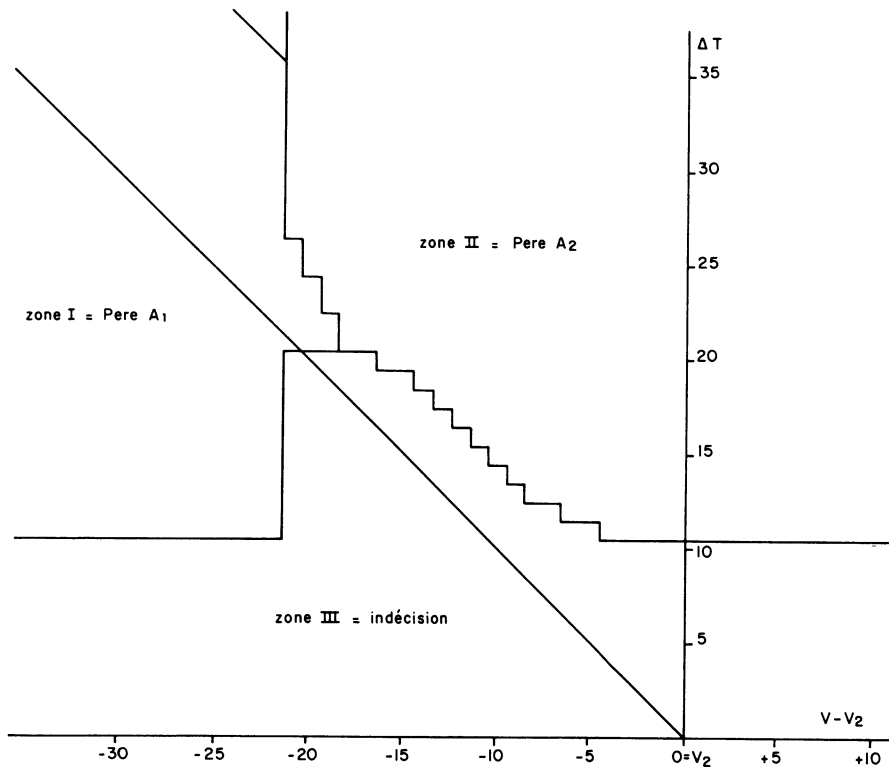
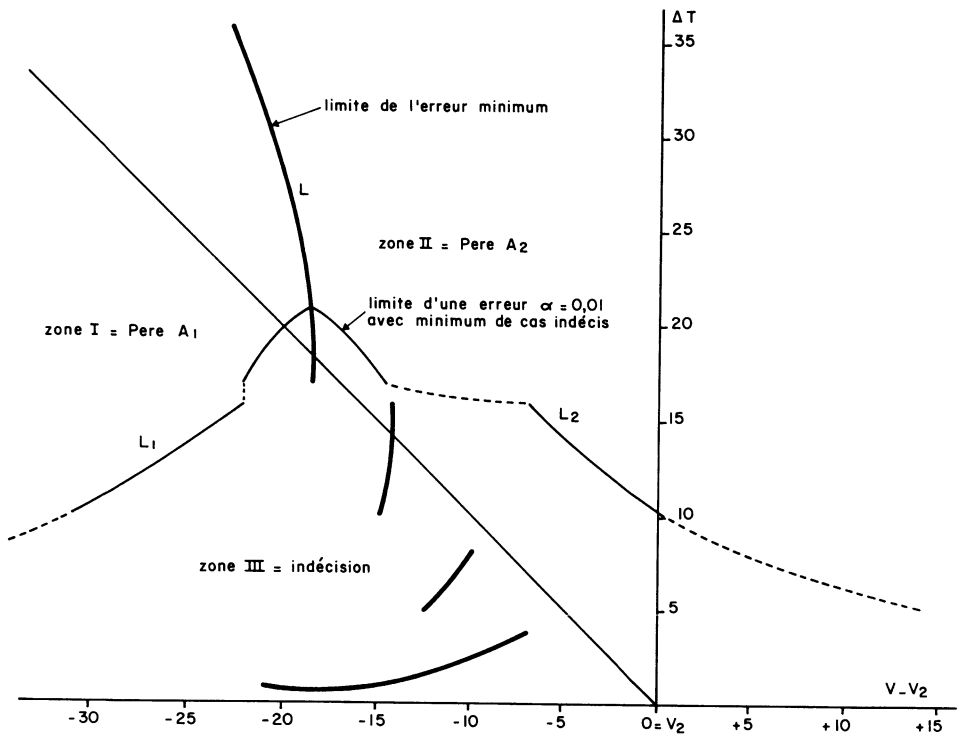


Figure 4

Avec analyse de groupe sanguin	Avec graphique utilisant la solution de Bayes
572 cas analysés 0 erreur	43 cas analysés 522 résolus graphiquement dont 7 erreurs (soit 1,22 %)

CONCLUSION

L'approche intuitive du problème par RENDEL (1958) permettait de résoudre une proportion non négligeable des cas de deux pères supposés. Toutefois, l'extension de l'insémination artificielle et de la diversité des taureaux, d'une part, la généralisation de l'état-civil, d'autre part, ont donné ces dernières années une grande importance à des sources d'erreurs jusqu'alors négligeables.

L'estimation, même approximative, de nouvelles limites a permis d'affiner considérablement les règles empiriques et de diviser par un facteur 10 le nombre des cas non résolus, tout en précisant le degré de vraisemblance des cas résolus.

La sélection repose sur une connaissance et une utilisation optimales des mécanismes de la transmission héréditaire. Cette connaissance est toujours de nature probabiliste : incertitudes dans l'estimation des aptitudes, doutes sur la généalogie, hasard des remaniements génétiques. A chacun des stades il est essentiel d'apporter le maximum de précision si l'on veut assurer la sécurité et l'efficacité des processus de sélection.

Figure 4 – Diagramme de détermination de paternité.

4a) Diagramme théorique.

Les résultats du tableau 2 permettent de construire un graphique situant les vélages observés par rapport aux dates prévues V_1 ou V_2 . L'origine des temps est fixée par V_2 (en jours).

L'écart entre vélages prévus est ΔT (en jours).

La ligne oblique figure les vélages V_1 .

Traits gras : limite L de séparation entre les pères.

Traits fins : limite L_1 et L_2 de la zone d'indécision (recours aux groupes sanguins).

4b) Diagramme pratique, simplifié.

De façon à ce que les zones comportent toujours un nombre entier de jours, les limites L , L_1 et L_2 sont dessinées en "marches d'escalier".

L'écart entre "vélage observé", V et "vélage prévu d'après la 2^e saillie", V_2 , est lu en abscisse.

L'écart entre saillie, donc entre vélages prévus, ΔT est lu en ordonnée.

Les zones I, II et III correspondent respectivement au premier père, au second père, ou à l'indécision.

BIBLIOGRAPHIE

- BOUGLER J. et DERVEAUX P., 1969. — Etude des durées de gestation dans l'espèce bovine, Union Nationale des Livres Généalogiques 16, rue Claude Bernard — 75531 Paris Cedex 05. 1-36.
- KENDALL M. and BUCKLAND W., 1960 — A Dictionary of Statistical Terms Olivier and Boyd. London. 22.
- *LEFEBVRE J., 1967. — Les groupes sanguins chez les animaux domestiques *Rev. d'immunol.*, 31, (1-2), 1-36.
- LHERMINIER P., 1972. — Normalisation des méthodes relatives à l'identification filiation des animaux de l'espèce bovine. Rapport I.T.E.B. 149 rue de Bercy 75579 Paris Cedex 12.
- LHERMINIER P., 1973. — Contrôles d'Etat-Civil par les groupes sanguins *Elevage & Insémination* 134, 17-26.
- RENDEL J. 1958. — Studies of cattle blood groups II Parentage test. *Acta agric. Scand.*, 8 : 131-215.
- RENDEL J., BOUW J. and SCHMID D.O., 1962. — The frequency of cows served twice which remain pregnant to first service : a study of results from parentage test. *Anim. Prod.* 4, 359-367.