

ESTIMATION DE LA DISTRIBUTION DES TAUX D'INFECTION INTRA CHEPTEL A PARTIR DES RESULTATS D'UN SONDAGE. PROPOSITION ET COMPARAISON DE DEUX METHODES*

Vincent Auvigne¹ et Etienne Petit²

RESUME

Deux méthodes, rassemblées dans un outil MS Excel dénommé DistriDiag, sont proposées pour estimer la distribution des taux d'infection intra-cheptel dans une population à partir des résultats d'un sondage. Les sondages pris en compte sont ceux réalisés par prélèvement d'un échantillon de taille fixe dans chaque cheptel testé. Ces méthodes sont appelées « modèle des centiles » et « modèle d'ajustement ascendant ». Il s'agit d'une application des lois de tirage au sort sans remise (loi binomiale, loi hypergéométrique). Les méthodes sont appliquées à un jeu de données réel provenant d'une enquête sérologique BVD réalisée avec des échantillons de cinq animaux dans 222 cheptels. Les deux méthodes aboutissent à des résultats similaires. Elles permettent de distinguer une population indemne ou faiblement infectée représentant 84 ou 85% des cheptels et une population fortement infectée représentant 16 ou 15% des cheptels. Cette formulation des résultats de ce sondage est notablement différente de celle qui serait classiquement donnée à la suite du calcul de l'intervalle de confiance : « 31 à 44% des cheptels testés sont infectés ». L'outil DistriDiag permet également de réaliser des simulations de stratégies d'échantillonnage (effet d'une modification de la taille de l'échantillon, du nombre de résultats positifs à partir duquel l'élevage est considéré positif). Les intérêts et limites de ces méthodes sont discutés.

Mots-clés : taux d'infection intra-cheptel, estimation, distribution.

SUMMARY

Two methods, brought together in an MS Excel tool called DistriDiag, are being proposed to estimate the distribution of intra-herd infection rates in a population, based on the results of a survey. The surveys concerned call for a fixed sample size in each herd tested. The methods are called « percentile method » and « ascending adjustment method ». They are based on the application of the laws of random sampling without replacement (binomial law, hypergeometric law). The methods were applied to a real set of data obtained from a serological BVD survey with samples of five animals in 222 herds. The two methods gave similar results. They make it possible to make a distinction between a population free of the disease or with a low infection rate, representing 84 or 85% of the herds, and a heavily infected population, representing 16 or 15% of the herds. The formulation of the results of the survey is notably different from that conventionally given following the calculation of a confidence interval: « 31 to 44% of the herds tested are infected ».

.../..

* Texte de la communication orale présentée au cours des Journées scientifiques AEEMA, 21 mai 2010

¹ Ekipaj, C/Valencia 3, 28223 Pozuelo de Alarcón, Espagne

² FRGDS Bourgogne, 42 rue de Mulhouse, 21000 Dijon, France

.../..

The DistriDiag tool also makes it possible to carry out strategic sampling simulations (the effect of a change in sample size, or of the number of positive results based on which a farm is considered positive). The advantages and limitations of the two methods are discussed.

Keywords: Intra-erd infection rate, Hypergeometric, Distribution.



I - INTRODUCTION

La fréquente prise en compte du troupeau comme unité épidémiologique est une caractéristique forte de l'épidémiologie vétérinaire. Elle la distingue de l'épidémiologie humaine, où les unités de regroupement de population sont rarement utilisées, à l'exception peut-être de leur utilisation dans les techniques de sondage en grappe [Rumeau-Rouquette *et al.*, 1985]. L'utilisation de l'unité troupeau amène à adapter les définitions des caractéristiques des tests de diagnostic. La sensibilité d'un test à l'échelle du troupeau est ainsi la probabilité d'obtenir au moins un résultat positif dans un cheptel infecté et la spécificité la probabilité de n'obtenir que des résultats négatifs [Toma *et al.*, 2001].

Une autre caractéristique importante de la stratégie de diagnostic au niveau du troupeau est la réalisation des tests sur un échantillon d'animaux. Ceci est particulièrement fréquent pour les productions porcine, avicole ou aquacole car la taille importante des cheptels rend difficile, pour des raisons pratiques et économiques, un prélèvement sur tout le cheptel. Dans ces productions, le prélèvement sur échantillon est la règle et le prélèvement exhaustif l'exception (cas du dépistage exhaustif des reproducteurs porcins en fin de prophylaxie de la maladie d'Aujeszky). En production bovine, les plans de contrôle sur échantillon sont moins fréquents du fait de la plus faible taille des cheptels et de la plus grande importance de la

médecine individuelle ; ils sont cependant utilisés dans certains cas comme des enquêtes de prévalence ou des protocoles établissant des statuts de troupeaux. Pour déterminer la taille des échantillons à prélever dans chaque troupeau, il est nécessaire de déterminer la prévalence intra-cheptel minimale que l'on souhaite détecter (ou la prévalence intra-cheptel supposée en cas d'infection d'un cheptel). Le nombre d'animaux à contrôler est ensuite calculé ou lu dans des tables [Toma *et al.*, 2001]. En pratique, comme le nombre d'animaux sur lesquels faire un prélèvement n'augmente plus à partir d'une certaine taille de cheptel³, les échantillons à prélever pour un plan de contrôle donné sont le plus souvent de taille fixe, ce qui simplifie la réalisation des opérations de terrain.

Les résultats d'un plan de contrôle d'un troupeau sur un échantillon des animaux sont donc un statut du troupeau (« positif » ou « négatif ») et, en cas de positivité, un pourcentage d'animaux « positifs » dans l'échantillon. Cette dernière information est cependant rarement exploitée. Une des raisons de la faible utilisation de cette information est probablement la difficulté de tirer des conclusions quantitatives d'un plan de contrôle qui a été conçu pour détecter l'infection et non la quantifier. Il est cependant regrettable qu'une information existante ne soit pas employée.

³ Si la prévalence limite intra-cheptel que l'on souhaite détecter est de 20% le nombre d'animaux à prélever est de 14 pour tout cheptel de taille égale ou supérieure à 200 animaux.

Dans ce contexte, l'objectif de cette communication est de présenter un cadre méthodologique et des outils pour l'interprétation de la distribution des prévalences intra-échantillons observées lors de la réalisation d'une enquête par échantillonnage d'animaux au sein de cheptels. On se limitera au cas où le

nombre d'animaux soumis à prélèvement est identique dans tous les cheptels. La présentation sera illustrée à partir d'un cas concret d'enquête sérologique dans une population d'élevages bovins. Ce cas concret est à considérer comme un jeu de données pris pour exemple.

II - MATERIELS ET METHODES

1. L'ENQUETE ET SES PRINCIPAUX RESULTATS

L'enquête prise en exemple est une enquête sérologique qui avait pour objectif d'objectiver l'importance de la circulation virale du BVD en Bourgogne [Petit, 2005]. Dans cette enquête, un prélèvement était effectué sur cinq génisses d'un an (8 à 16 mois) dans un échantillon d'élevages tirés au sort dans les quatre petites zones naturelles de la région.

Dans l'hypothèse d'une taille moyenne de 50 animaux par élevage, l'échantillon d'animaux soumis à un prélèvement permettait de détecter, avec un risque d'erreur par défaut inférieur à 5%, une prévalence minimale de 50%. Les prélèvements étaient analysés par une Elisa sur sérum individuel.

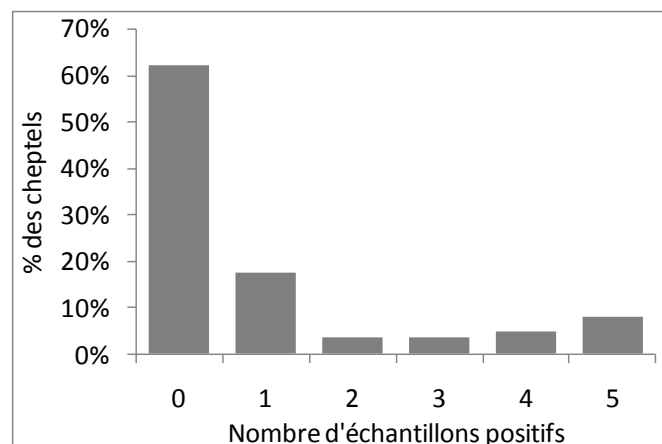
Deux cent vingt deux cheptels ont été étudiés et 84 se sont avérés « positifs » (au moins un prélèvement positif). Le taux d'infection des cheptels dans cet échantillon de cheptels est donc de 38%. Un intervalle de confiance à 95%

peut être calculé [31 à 44%] (Epidat 3.1). Les résultats peuvent donc être formalisés de la façon suivante « 31 à 44% des cheptels testés sont infectés ». Certains pourront cependant préférer la formulation « 56 à 69% de cheptels sont infectés à moins de 50% ou sont indemnes ».

Cependant, aucune de ces deux formulations ne prend en compte le pourcentage observé d'animaux infectés au sein des cheptels infectés. Il s'agit cependant d'une information dont l'on dispose (figure 1). La distribution des résultats observée dans ce cas concret a ses caractéristiques particulières. On note en particulier une distribution « en U » : les modalités les plus fréquentes sont les échantillons négatifs et faiblement positifs (un positif sur cinq), d'une part, et fortement positifs, d'autre part (cinq positifs sur cinq). Les cas intermédiaires sont moins fréquents. La question qui se pose alors est s'il est possible de prendre en compte cette information dans l'analyse des résultats, et de quelle manière.

Figure 1

Enquête BVD - Distribution observée du nombre d'échantillons « positifs » au sein des élevages à réponse(s) positive(s)



2. LES CONCEPTS DE PROBABILITES UTILISES DANS LE MODELE

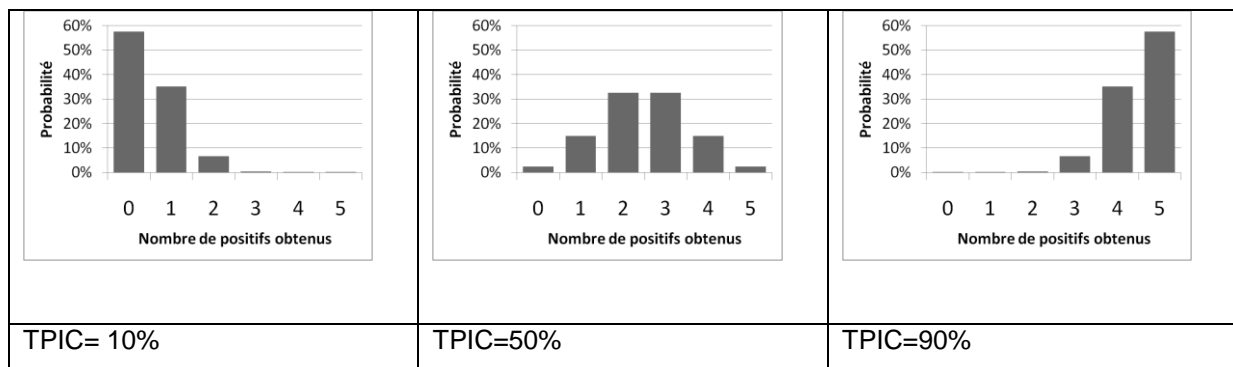
Un élément important à prendre en compte est que la distribution observée des nombre de « positifs » n'est pas une image directe de la distribution des prévalences intra-cheptels dans la population étudiée. En effet, les lois de la probabilité interviennent. Ceci est illustré par la figure 2 où sont représentées pour différentes situations de taux de prévalence intra-cheptel (TPIC) les probabilités d'obtenir les différents résultats possibles sur un échantillon de cinq animaux prélevés au hasard sur 50. Il s'agit d'une modélisation d'un tirage sans remise par une loi hypergéométrique [Pouillot, 2002]. On constate que si le taux de prévalence intra-cheptel est de 50%, dans 2,5% des cas aucun animal « positif » ne sera détecté ; on retrouve là, exprimé d'une autre manière, le fait que l'échantillon de cinq animaux permet de détecter une prévalence intra-cheptel de 50% avec un risque d'erreur par défaut inférieur à 5%. A contrario, si la prévalence intra-cheptel n'est que

de 10%, la probabilité de ne pas détecter de cheptel « positif » devient très élevée (58%) ; ceci nous confirme qu'un échantillon de cinq n'est pas adapté à détecter une telle prévalence. Cependant, il faut aussi remarquer que dans un nombre non négligeables de cas (42%) l'infection sera détectée. Le fait que la taille de l'échantillon n'est pas suffisante pour détecter un taux de prévalence avec une bonne sécurité ne signifie pas que l'infection ne sera jamais détectée !

La distribution observée des nombres de positifs dans l'échantillon dépend donc de deux éléments : d'une part de la distribution réelle des prévalences intra-cheptels dans la population observée, qui est inconnue, et d'autre part des lois du tirage au sort sans remise dans une population finie telles que rappelées ci-dessus, qui sont connues. L'objectif de la méthodologie présentée ci-après est de déterminer, à partir des éléments connus, la *distribution vraisemblable* des taux de prévalence intra-cheptel.

Figure 2

Distribution des probabilités d'obtention d'un résultat d'un échantillon de cinq individus tirés au sort dans une population (troupeau) de 50. TPIC = Taux de prévalence intra-cheptel

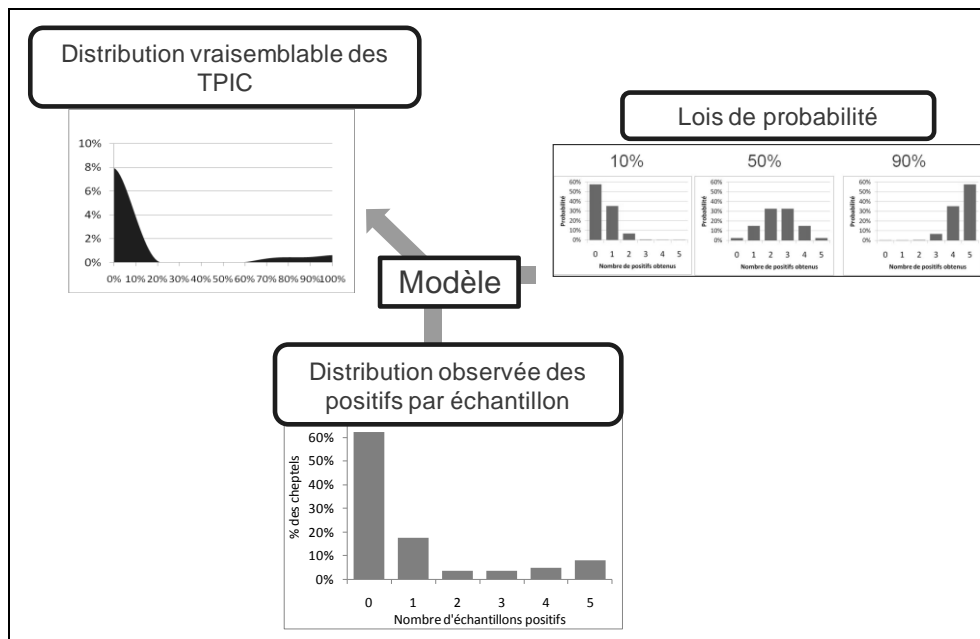


3. PRINCIPES DES MODELES

Deux modèles sont proposés pour estimer la *distribution vraisemblable* des taux de prévalence intra-cheptel à partir de la distribution observée des nombres de positifs dans les échantillons.

Les deux utilisent des méthodes probabilistes, sont développés sur Microsoft Excel® et rassemblés dans un outil dénommé **DistriDiag**. Les deux modèles ont le même objectif : rechercher la ou les distributions vraisemblables des taux de prévalence intra-cheptel (figure 3).

Figure 3
Principes de la modélisation



3.1. MODELE DES CENTILES

Ce modèle estime les centiles de la distribution vraisemblable des TPIC. Pour ce faire, le modèle cherche à minimiser, à l'aide de la fonction *solveur* d'Excel, une variable V . Cette variable est la somme de :

- La somme (a) des écarts au carré entre la distribution observée et la distribution calculée obtenue par application des lois du tirage au sort sur une population de *distribution vraisemblable*. (a) est égal à 0 si la distribution calculée est égale à la distribution observée ;
- La somme (b) des écarts aux carrés entre deux centiles successifs de la distribution vraisemblable, multiplié par un coefficient (c).

Le coefficient (c) est appelé *coefficient d'autocorrélation*. L'autocorrélation est la corrélation qui existe entre les valeurs successives d'une variable. Il s'agit d'un concept classiquement utilisé pour des variables temporelles (la température d'aujourd'hui n'est pas indépendante de celle de demain) [Toma *et al.*, 1991], ou spatiales (un taux de prévalence dans une zone n'est pas indépendant des zones voisines) [Estève *et al.*, 1993]. Ce concept est utilisé ici pour définir le fait qu'il existe une certaine corrélation entre les valeurs de deux centiles successifs. Le coefficient

d'autocorrélation permet de moduler l'importance donnée à ce phénomène et limiter les degrés de liberté du modèle. Plus il est élevé plus la distribution vraisemblable estimée sera lissée et plus le coefficient (a) sera élevé. La valeur de (c) est déterminée lors de chaque analyse de distribution par avis d'expert. Il s'agit donc de trouver un compromis entre, d'une part la reproduction de la distribution observée des nombres de positifs par échantillon et, d'autre part, la compatibilité entre la distribution vraisemblable des TPIC modélisée et les hypothèses sur l'histoire naturelle de la maladie étudiée.

3.2. MODELE D'AJUSTEMENT ASCENDANT

Ce modèle vise à représenter la distribution vraisemblable des TPIC dans la population par un nombre limité de sous-populations à TPIC fixe (sans variabilité intra sous-population). L'objectif est de réaliser une modélisation avec le minimum de sous-populations. Ce modèle est appelé « d'ajustement ascendant » car des itérations sont réalisées en commençant par un modèle avec une seule sous-population puis en augmentant progressivement le nombre de sous-populations jusqu'à l'obtention d'un ajustement satisfaisant. Lors de chaque itération, le modèle calcule à partir de la loi de probabilité (binomiale

ou hypergéométrique lorsque la taille de l'échantillon est importante par rapport à la taille de l'élevage) la distribution théorique observée avec l'échantillonnage effectué. Le solveur cherche à ajuster la proportion de chaque sous-population ainsi que son TPIC respectif pour minimiser la distance du Chi² entre la distribution

théorique et la distribution observée des échantillons. Si la distance du Chi² est trop élevée ($p(\text{Chi}^2) < 0,2$), on réalise une nouvelle itération après avoir ajouté une sous-population. Des itérations successives sont réalisées jusqu'à obtention d'une distance jugée satisfaisante.

III - RESULTATS

1. ESTIMATION DE LA DISTRIBUTION VRAISEMBLABLE

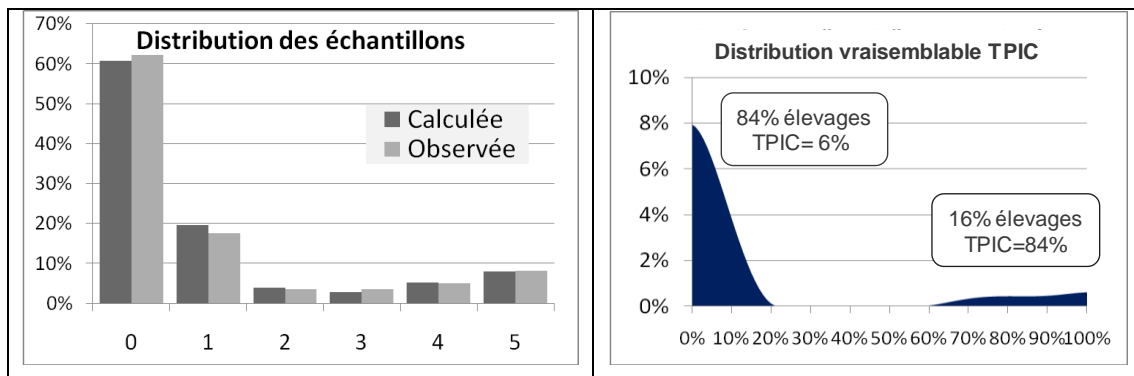
1.1. ESTIMATION AVEC LE MODELE DES CENTILES

Les résultats obtenus sont présentés figure 4. L'histogramme de gauche permet de comparer les distributions observée et calculée. La courbe en aire de droite donne la distribution

vraisemblable des TPIC dans la population d'origine. Dans ce cas concret d'enquête BVD, la distribution vraisemblable est nettement bimodale avec deux populations disjointes : une population indemne ou faiblement infectée représentant 84% des cheptels et une population fortement infectée représentant 16% des cheptels.

Figure 4

Modèle des centiles - Estimation de la distribution vraisemblable



1.2. ESTIMATION AVEC LE MODELE D'AJUSTEMENT ASCENDANT

Dans une première itération un modèle à une seule sous-population est testé. Avec ce modèle, les distributions théoriques et observées des échantillons sont différentes (figure 6, Chi², $p < 0,01$). Ce modèle est donc rejeté et un modèle à deux sous-populations est testé. Avec ce modèle à deux sous-populations les distributions théoriques et observées des échantillons ne sont pas différentes (figure 6, Chi², $p = 0,86$). Ce modèle est donc retenu. Les caractéristiques des deux sous-populations sont décrites dans le tableau 1.

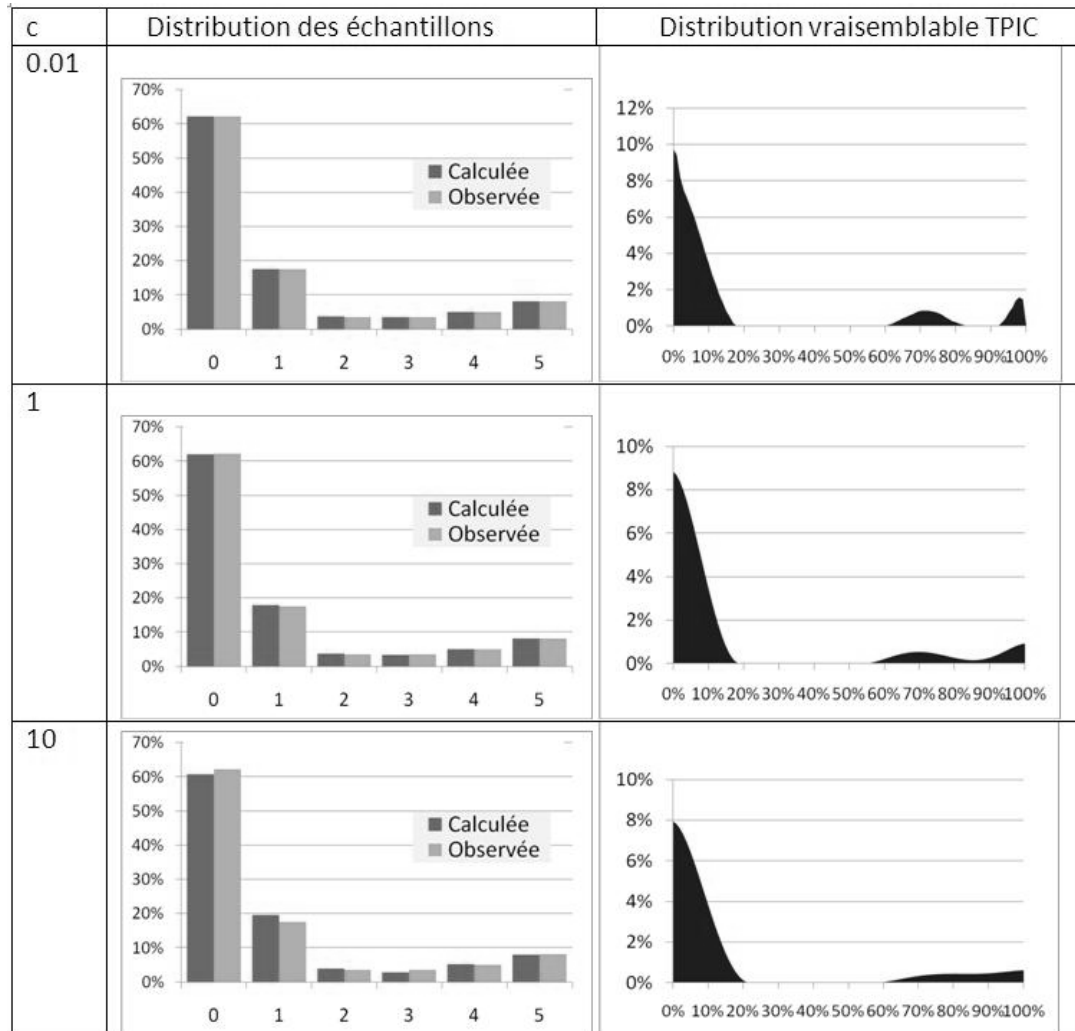
2. EXEMPLES D'APPLICATIONS PRATIQUES

Les deux méthodes proposées sont des outils d'aide à l'observation des résultats. En permettant de discriminer des sous-populations elles facilitent la réflexion sur l'histoire naturelle de la maladie. Dans le cas concret étudié, il est probable que les deux sous-populations correspondent à des phénomènes biologiques complètement différents. L'hypothèse est que la population fortement infectée correspond à des élevages ayant subi une circulation virale massive et que la population faiblement infectée correspond soit à des résidus d'anticorps maternels soit à des défauts de spécificité des tests.

Figure 5

Permet de visualiser l'effet du coefficient d'autocorrélation.

S'il est faible (0,01) la correspondance entre les distributions calculées et observées est très bonne mais les élevages sont regroupés en trois sous-populations. S'il est élevé (10), les distributions calculées et observées divergent légèrement et il n'y a que deux sous-populations. Le coefficient fort a été retenu.



En lien avec le point précédent, les méthodes proposées peuvent aussi contribuer à la validation des stratégies d'échantillonnage. Dans le cas concret étudié, il est clair que si la population faiblement infectée est d'importance épidémiologique, l'échantillon de cinq est inadapté pour différencier les faiblement infectés des indemnes. A l'opposé, si on considère que

les faiblement infectés ne font pas courir de risque épidémiologique il faut prendre en compte que certains de ces élevages seront détectés, bien que le sondage ait été « conçu pour détecter les élevages infectés ». Ne faudrait-il pas alors ne considérer comme « positifs » que les élevages ayant plusieurs individus « positifs » ?

Figure 6

Modèle d'ajustement ascendant.

Distribution des échantillons théoriques et observés suivant le nombre de sous-populations.

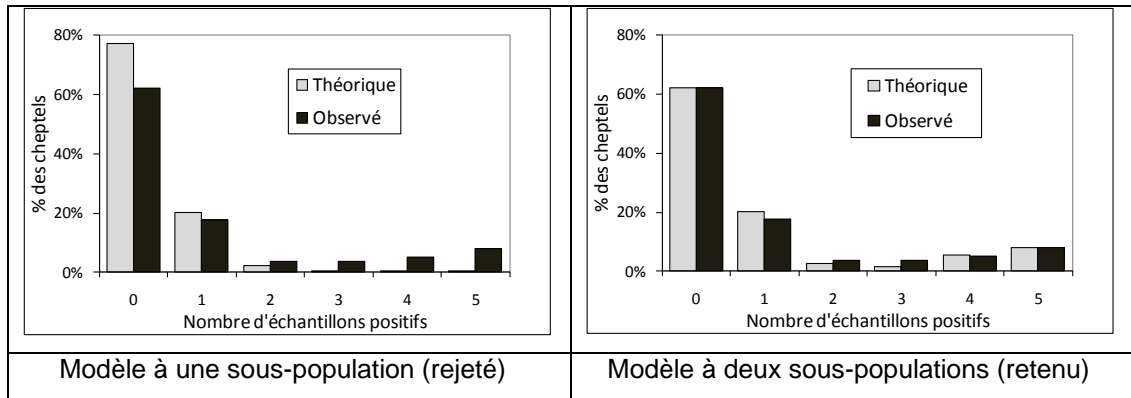


Tableau 1

Modèle d'ajustement ascendant.

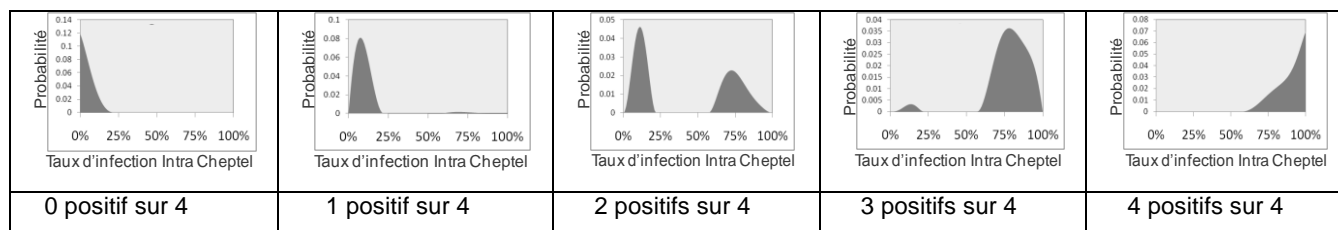
Caractéristiques des deux sous-populations du modèle retenu.

	% des cheptels	TPIC [IC 95%]
Population 1	85%	6,0% [4,5 – 7,6]
Population 2	15%	88,4% [83,0 – 93,8]

Pour répondre à ces questions, DistriDiag offre des outils pour le praticien et l'épidémiologiste afin d'aider à définir les règles de décision qui permettent de classer un élevage donné dans l'une ou l'autre des sous-populations vraisemblables. C'est une approche bayésienne [Pouillot et Sanaa, 2002] car on estime que la probabilité qu'un élevage appartienne à telle ou telle sous-population dépend, en plus du résultat des tests réalisés dans l'élevage, d'une information *a priori* : la distribution vraisemblable des TPIC dans la population dont est issu l'élevage prélevé. Il s'agit en fait d'un concept proche des valeurs prédictives positives et négatives : on utilise l'information disponible sur la distribution du pathogène dans la population pour interpréter le résultat du test de diagnostic. L'outil permet de faire des simulations pour des

nombre d'animaux prélevés différents de celui de l'étude ayant permis de déterminer la distribution vraisemblable des TPIC. Nous présentons ici une simulation d'un échantillonnage de quatre animaux par cheptel (figure 7). Dans le cas concret étudié, on peut conclure que s'il y a zéro ou un échantillon « positif » sur quatre, l'élevage est à classer dans la catégorie « indemne ou faiblement infecté », que s'il y a trois ou quatre « positifs », l'élevage est fortement infecté, et que s'il y a deux positifs il est impossible de trancher (52% de chances d'être faiblement infecté contre 48% d'être fortement infecté). Heureusement, dans le cas étudié, cette configuration (deux positifs) est rare (4% des cas). Il est aussi important de noter qu'il est impossible de dire si un échantillon avec zéro positif provient d'un élevage indemne ou d'un élevage faiblement infecté.

Figure 7
Simulation d'un échantillonnage de quatre animaux par cheptel : probabilité de TPIC en fonction du nombre de positifs dans l'échantillon



IV - CONCLUSION - DISCUSSION

Les méthodes proposées utilisent une information souvent sous-exploitée lors des enquêtes par échantillonnage au niveau du troupeau : la distribution du nombre de « positifs » par troupeau. Elles peuvent également être utilisées si l'unité de regroupement n'est pas un troupeau mais une bande, un lot... Dans le cas concret étudié, elles aboutissent à une formulation des résultats nettement différente de celle de méthodes classiques de calcul : alors que la méthode classique permet d'affirmer que le taux d'infection des cheptels est de 38%, les méthodes proposées aboutissent à affirmer que 15 ou 16% (suivant la méthode utilisée) des élevages sont fortement infectés. La différence entre les deux méthodes est loin d'être négligeable. Doit-elle amener les épidémiologistes à modifier la formulation des résultats de ces types de sondage ?

Un des intérêts majeurs de ces méthodes est de fournir un cadre méthodologique pour évaluer *a posteriori* la pertinence d'une stratégie d'échantillonnage. En effet, dans les approches classiques, il est nécessaire de faire des hypothèses sur les TPIC pour déterminer la taille de l'échantillon à prélever dans chaque troupeau mais, à notre connaissance, aucune méthode n'est proposée pour évaluer *a posteriori* la pertinence des choix effectués. Un autre intérêt de l'approche proposée est de centrer l'observation et l'analyse sur l'existence de sous-populations de cheptels, caractérisées par des TPIC différents. Cette « mise en classe » peut faciliter la réflexion sur les phénomènes biologiques mis en œuvre. Cette approche par « mise en classe » serait à comparer avec celles utilisées dans les méthodes d'appréciation probabiliste quantitative du risque où on cherche

à modéliser le TPIC, entre autres paramètres, par une loi de probabilité continue [Pouillot *et al.*, 2002].

Le cas du sondage BVD utilisé pour cet exposé est particulièrement net. Les deux sous-populations obtenues sont nettement distinctes ; ceci se traduit, quand *le modèle des centiles* est utilisé, par une faible sensibilité des résultats à la modification du coefficient d'autocorrélation et, quand le modèle *d'ajustement ascendant* est utilisé, par un ajustement satisfaisant avec seulement deux sous-populations. Les méthodes ont été testées avec d'autres jeux de donnée (résultats non présentés). Les résultats sont parfois plus confus avec plus de difficultés à choisir le coefficient d'autocorrélation (M. des centiles) ou le nombre de sous-populations (M. d'ajustement ascendant). Ceci ne met pas en cause le principe des méthodes proposées mais provient probablement du fait que, selon les processus biologiques observés, les situations sont plus ou moins tranchées. Une hypothèse est qu'il y a une certaine dichotomie entre des processus de type viral, où les situations sont très tranchées, et des processus de type bactérien où la tendance serait plus à un gradient de TPIC sans sous-populations nettement distinctes.

Deux méthodes sont proposées. Chacune a été développée par l'un des auteurs. Elles sont fondées sur des principes similaires et, à ce jour, aucun élément décisif ne permet de préférer l'une ou l'autre. Les résultats obtenus avec les deux méthodes sont similaires. Le *modèle des centiles* permet de donner un élément d'appréciation de la variabilité des TPIC au sein des différentes sous-populations et n'oblige pas à fixer d'hypothèses sur le nombre de sous-populations. Elle a cependant comme

inconvenient de nécessiter la mise en œuvre d'un coefficient d'autocorrélation dont la nature n'est pas évidente à percevoir, ce qui peut nuire à la confiance accordée aux résultats ; il peut paraître aussi curieux que le nombre de paramètres à estimer (101 centiles) soit supérieur au nombre de classes observées. La méthode *d'ajustement ascendant*, en minimisant le nombre de sous-populations et en prenant comme hypothèse l'absence de variabilité des TPIC au sein de chaque population, aboutit à des résultats plus simples à percevoir et à exprimer. Son inconvenient est de ne pas donner d'éléments pour apprécier si deux sous-populations sont nettement distinctes ou plus entremêlées. Il sera nécessaire d'utiliser ces deux méthodes dans d'autres cas concrets pour choisir entre les deux, si un choix était à faire.

Ces outils ont leurs limites. Ils nécessitent en particulier un certain nombre de simplifications. Il faut que tous les échantillons soient de taille identique, la taille du troupeau est fixée et le résultat de l'analyse doit être individuel et qualitatif (binaire). Dans l'état de développement actuel des outils, si tous les échantillons ne sont pas de taille identique il faut soit les exclure, soit les ramener à la taille standard par une règle de trois. En ce qui concerne la problématique de la taille du troupeau, on peut estimer que l'impact est faible si les troupeaux sont assez grands.

Une autre limite importante est l'estimation de l'intervalle de confiance des résultats obtenus. La méthode des centiles n'en fournit aucun. La méthode d'ajustement, fondée sur le maximum de vraisemblance, permet un calcul d'intervalle de confiance des taux de prévalence intra cheptel au sein de chaque sous-population, à condition de ne pas utiliser un modèle saturé, mais ne donne pas d'intervalle de confiance de l'importance relative des différentes sous-populations. Une piste serait d'utiliser des méthodes de type « Bootstrap ». Elle reste à explorer.

Une autre limite, particulière au *modèle des centiles*, est qu'aucun critère d'ajustement n'est proposé pour choisir le modèle optimal : le choix du coefficient d'autocorrélation est laissé à avis d'expert. Mais ceci est-il une limite ? N'est-il pas

illusoire de vouloir se reposer sur un paramètre pour faire un choix ? N'est-il pas préférable d'assumer qu'il n'y a pas un modèle unique pour représenter la réalité ? L'épidémiologiste peut aussi revendiquer une part d'art vétérinaire dans son travail...

La méthode proposée ne doit pas être confondue avec les méthodes « mixtures of distributions » qui sont utilisées pour décrire des sous-populations à partir d'une distribution d'un caractère quantitatif ; ces méthodes peuvent par exemple être utilisées pour affiner les critères d'interprétation d'une réaction à la tuberculine [Shashidhar *et al.*, 2002]. Nos méthodes consistent simplement à « appliquer un filtre » pour prendre en compte le « bruit » apporté par les lois de probabilité lors de la réalisation de l'échantillonnage. Les deux techniques pourraient être complémentaires : une méthode « mixtures of distributions » pourrait être utilisée pour aller plus loin sur l'analyse de la distribution vraisemblable donnée par la méthode des centiles. Ceci n'aurait pas d'intérêt dans le cas concret présenté dans cet article car les deux sous-populations vraisemblables sont nettement distinctes. Ceci n'est cependant pas toujours le cas : quand il n'y a pas de cut-off net entre les sous-populations (anti-mode pas net), les méthodes « mixtures of distributions » pourraient être un complément intéressant.

Enfin, les outils de calcul ne permettent pas de répondre à toutes les questions. Une des motivations à l'origine de ce travail était d'arriver à déterminer, grâce à l'analyse des distributions, quel était le pourcentage de troupeaux vraiment indemnes, sachant que s'il y a une sous-population faiblement infectée certains élevages infectés n'ont pas été détectés. La conclusion, provisoire peut-être, est que ceci est impossible à déterminer : on arrive uniquement à estimer la fraction de population faiblement infectée ou indemne, sans pouvoir les distinguer. On peut néanmoins, dans la méthode ascendante, imposer au modèle la présence d'une sous-population de prévalence nulle et estimer sa proportion, mais cela amène souvent à créer une sous population supplémentaire. L'analyse des chiffres ne peut pas tout, il faudra retourner sur le terrain avec une autre stratégie d'échantillonnage si on veut approfondir cette question...

BIBLIOGRAPHIE

- Estève J., Benhamou E., Raymond L. - Méthodes statistiques en épidémiologie descriptive. Les éditions INSERM, Paris, France, 1993. 307 p.
- Petit E. - Mise en place d'un observatoire régional de la maladie des muqueuses (BVD) en Bourgogne. *Epidémiol. et santé anim.*, 2005, **48**, 1-8.
- Pouillot R. - Caractérisation d'une loi de distribution d'une variable entrant dans un modèle d'analyse de risque probabiliste. *Epidémiol. et santé anim.*, 2002, **41**, 113-143.
- Pouillot R., Sanaa M. - Bases probabilistes et statistiques nécessaires à l'analyse du risque. *Epidémiol. et santé anim.*, 2002, **41**, 67-83.
- Pouillot R., Sanaa M., Dufour B. - Principes de l'appréciation probabiliste quantitative des risques. *Epidémiol. et santé anim.*, 2002, **41**, 94-112.
- Rumeau-Rouquette C., Bréart G., Padiou R. - Méthodes en épidémiologie. 3^{ème} éd., Médecine-Sciences Flammarion, Paris, 1985. 398 p.
- Shashidhar J., Savanur V., Chadha, Jagannatha P. - Mixture model for analysis of tuberculin surveys. *Ind. J Tub.*, 2002, **49**, 147-152.
- Toma B., Bénét J-J., Dufour B., Eloit M., Moutou F., Sanaa M. - Glossaire d'épidémiologie animale. Editions du Point Vétérinaire, Maisons-Alfort, 1991. 365 p.
- Toma B., Dufour B., Sanaa M., Bénét J-J., Shaw A., Moutou F., Louza A. - Épidémiologie appliquée à la lutte collective contre les maladies transmissibles majeures. 2^{ème} éd., AEEMA Eds, Maisons-Alfort, 2001. 690 p.



Remerciements

Les auteurs remercient Régis Pouillot et Benoît Durand pour leur contribution essentielle aux premières étapes de réflexion autour de ce projet et Christelle Fablet, Nicolas Rose et Benoît Gouvars pour leurs apports à la discussion des résultats.