

prevalence as set by the null hypothesis. This cumulative probability is compared with α to assess whether the sample size is correct, too large or too small.

ANALYSIS OF SURVEY RESULTS

Analysis of survey results is a much simpler probability calculation. The same parameters as used for sample size calculation are required, in addition to the actual sample size used and the number of reactors observed. The program then reports the probability of observing this number of reactors under the null hypothesis, and provides an interpretation as to whether the null can be rejected or not.

TWO-STAGE SAMPLING

Two-stage sampling for surveys of large populations is necessary, both because of the clustered nature of disease, and from a practical point of view (eg. the problem of generating a comprehensive sampling frame). The first stage involves sampling herds (or any other convenient grouping of animals), and the second, individual animals from the herd. A two stage analysis approach must be used, first to classify each herd sampled as infected or uninfected based on the results of the individual animal tests, and then to classify the population of herds as infected or uninfected, based on the herd or aggregate tests.

A screening test for an individual animal is characterised by its sensitivity and specificity. When a screening test is applied to a sample of animals from a herd for the purpose of classifying the herd as diseased or non-diseased, the combined procedure may be thought of as a single, herd-level screening test, with its own sensitivity and specificity. The sensitivity and specificity of herd tests are influenced by the sensitivity and specificity of the individual animal test used, the number of animals tested, as well as the way in which individual animal results are interpreted. Various authors have discussed the interpretation of herd tests (Martin et al., 1992; Donald et al., 1994; Jordan, 1995). The most common approach is to use a cut-point number of reactors to classify the herd as either diseased or non-diseased. If the cut-point chosen is zero, herd-level specificity will be low, but the sensitivity will be high. As the cut-point number of reactors is increased the sensitivity decreases and the specificity increases. Designing an optimal herd-level survey is therefore a question of determining the best sample size and the cut-point level, that will give the optimal combination of sensitivity and specificity for the purposes of the survey. Ideally, the researcher should be able to determine the herd test sensitivity and specificity required and then calculate the corresponding sample size and cut-point number of reactors.

To calculate the sample size for single-stage surveys, it is necessary to first define the levels of Type I and II errors. In the context of herd testing, the herd-level sensitivity is the probability that a diseased herd will be classified as diseased. This is equal to one minus the probability that a diseased herd will be classified as disease-free, or $1 - \alpha$. Similarly, the herd-level specificity is equal to $1 - \beta$. These simple relationships allow us to specify the required herd-level sensitivity and specificity by specifying the animal-level power and confidence levels, and calculate the animal-level sample size and cut-point number of reactors needed to achieve them.

DISCUSSION

The FreeCalc program is able to carry out all the calculations necessary for the determination of sample size and the analysis of surveys to prove freedom from disease. The program allows field officers with little statistical training to conduct precisely designed surveys easily and with confidence. The flexibility of the program also offers the opportunity to calculate least-cost sample sizes for two-stage sample surveys. The program is available free of charge over the Internet on the World Wide Web at the EpiVetNet Web site (<http://epiweb.massey.ac.nz>). No restriction is placed on the distribution of the program, and users are encouraged to pass it on to colleagues. Use of the program should always be acknowledged in reports, scientific papers and presentations.

ACKNOWLEDGEMENTS

The formula and program described in this paper were developed under a project funded by the Australian Centre for Agricultural Research and working in cooperation with the Thai Department of Livestock Development. Angus Cameron is supported by a Junior Research Fellowship from the Australian Meat Research Corporation.

BIBLIOGRAPHY

- Anon. 1994. General agreement on tariffs and trade (GATT), Sanitary & Phytosanitary Measures (MTN/FA II-A1A-4). Agreement on the application of sanitary and phytosanitary measures
- Donald A.W., Gardner I.A., Wiggins A.D., 1994. Cut-off points for aggregate herd testing in the presence of disease clustering and correlation of test errors. *Preventive Veterinary Medicine* 19, 167-187
- Jordan D. 1995, Aggregate testing for the evaluation of Johne's disease herd status. In: Morton J. (ed) *Epidemiology Chapter, Australian College of Veterinary Scientists Proceedings. Australian Veterinary Association Annual Conference, Melbourne, May 21-26, 1995. Australian College of Veterinary Scientists, Indooroopilly, 60-67*
- Martin S.W., Shoukri M., Thorburn M.A., 1992. Evaluation the health status of herds based on tests applied to individuals. *Preventive Veterinary Medicine* 14, 33-43

UN OUTIL PARAMÉTRIQUE DE CLASSIFICATION QUI PEUT ESTIMER SIMULTANÉMENT SENSIBILITÉ, SPÉCIFICITÉ ET PRÉVALENCE : LA DISTRIBUTION MIXTE

Bigras-Poulin M.¹

Diagnostic is in itself a classification problem. Interpretation and medical decision making while using a serological diagnostic test is dependent upon the use of a cut-off value. The choice of this value is dependent upon the characteristics of the underlying reaction of the individuals from the healthy and the diseased populations to the test. We suggest that making the assumption that the test results from both of these populations are normally distributed with $N(\mu_d, \sigma_d^2)$ and $N(\mu_h, \sigma_h^2)$. The general population is composed of individuals from two distinct sub-populations, the healthy and the diseased. The prevalence (p) indicates the proportion of individuals from the diseased sub-population which are present in the general population. This situation can be modeled by the density function of a mixture of two normal populations. A deterministic simulation study was undertaken to better understand the difficulty that occur when estimating the five parameters ($p, \mu_d, \sigma_d^2, \mu_h, \sigma_h^2$) of a mixture of two normal distributions. It was found that, in certain situations, a large sample was needed to correctly estimate these parameters but, in difficult cases, the poor results are likely to be of little practical importance. In most cases, a sample size of 200 to 300 is sufficient and is usually available. Some authors have shown that this approach is very useful and can be enhanced by the addition of a sample of individuals for which the membership to the healthy or diseased group is known.

INTRODUCTION

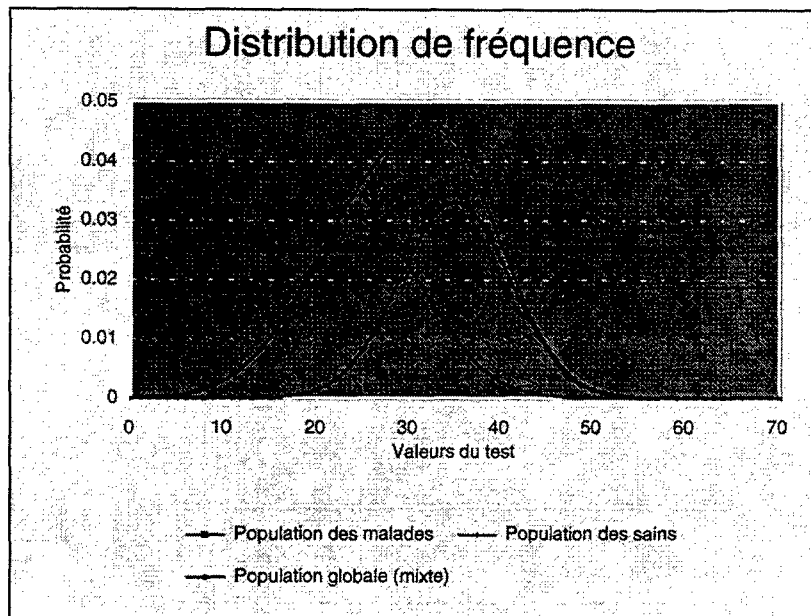
Le diagnostic fait partie de la problématique de la classification dans la décision médicale car il faut décider si l'animal ou le troupeau sont malades/infectés ou sains. Dans le cas d'un diagnostic sérologique, la décision est souvent prise en utilisant comme seule donnée le résultat de la sérologie. Dans la suite de la présentation nous utilisons le mot malade pour représenter indifféremment malade ou infecté car, pour l'objet de la présentation, seul le test utilisé déterminera s'il s'agit d'une recherche de malades ou d'infectés. La méthode ELISA est un outil de diagnostic souvent utilisé. Cette méthode transmet un résultat sous la forme d'une densité optique que l'on peut considérer comme une variable continue. Il est raisonnable de présumer qu'une population homogène par rapport à la maladie ou l'infection présentera des résultats de sérologie qui suivront une distribution gaussienne. Ceci nous permet alors de considérer que la population des individus sains soumis au test sérologique nous fournira une variable aléatoire distribuée normalement selon $N(\mu, \sigma^2)$. La population des individus malades nous fournira un résultat comparable distribué normalement selon $N(\mu, \sigma^2)$.

Dans une situation diagnostique habituelle, le statut des individus par rapport à la maladie est inconnu. La population est alors constituée d'un mélange d'individus sains et d'individus malades. La prévalence décrit la proportion d'individus malades dans la population. Celle-ci peut donc être considérée comme le résultat d'un mélange d'une population d'individus sains et d'une population d'individus malades selon une proportion indiquée par la prévalence. La distribution de fréquence de cette dernière population correspondra donc à un mélange de deux distributions normales aussi nommée distribution mixte (Figure 1). Une distribution mixte constituée d'un mélange de deux normales aura cinq paramètres, soit la proportion, les deux moyennes et les deux variances.

La distribution mixte n'est pas une nouveauté en statistique car Pearson (1894) l'a étudiée ainsi que Charlier (1906). Ces auteurs ont abordé l'étude de ces distributions en utilisant l'approche des moments. Hasselblad (1966) propose l'utilisation de la vraisemblance maximale pour estimer les paramètres ainsi que Day (1969), Wolfe (1970) et Hosmer (1972). Le principal problème indiqué par ces auteurs, pour le cas univarié, est la nécessité d'un échantillon de taille importante. Dans ce cas, Hathaway (1985) a montré que l'utilisation du maximum local comme estimateur est consistant et efficace. Cette approche présente cependant des problèmes quand à l'estimation des paramètres de la distribution mais les avantages de son utilisation sont importants. Une comparaison de différentes méthodes d'estimation a montré que la vraisemblance maximale accompagné de l'algorithme de Newton est une méthode performante dans le cas univarié (Everitt, 1984). Titterington et al (1985) ont publié un texte de référence complet sur l'analyse statistique des distributions mixtes.

¹ Université de Montréal, Faculté de Médecine Vétérinaire, Département de pathologie et microbiologie, 3200 Sicotte, Saint-Hyacinthe (Québec), Canada, J2S 7C6

Figure 1
Exemple de distribution mixte composée de deux normales (prévalence 40%)



Hosmer (1973) a étudié l'amélioration que l'on peut obtenir lors de l'estimation des paramètres si l'on connaît l'appartenance d'une partie des individus sous observation à la sous-population des sains et à celle des malades. Cette amélioration est souvent possible car pour certains individus nous avons un supplément d'observations cliniques ou autres permettant d'avoir une certaine assurance quant à l'appartenance des individus à une ou l'autre des sous-populations.

La sensibilité et la spécificité sont des mesures d'efficacité des tests qui ne sont pas généralement disponibles. De plus, quand elles sont disponibles, elles sont estimées dans des conditions qui sont souvent peu représentatives des conditions rencontrées dans le fonctionnement habituel des laboratoires de diagnostic. Par exemple, le laboratoire de diagnostic est soumis à des contraintes économiques qui font qu'un test n'est pas répété plusieurs fois pour s'assurer de la qualité du résultat, ce qui est souvent fait par un laboratoire de recherche qui est soumis à des contraintes économiques bien différentes. Lors des études servant à estimer sensibilité et spécificité, les populations de références sont souvent obtenues par des infections expérimentales et par une population d'individus sains ayant vécu dans un environnement grandement protégé. Ceci fait que la valeur estimée, quand elle est disponible, a tendance à surestimer les valeurs des mesures d'efficacité du test. Il est donc souhaitable d'obtenir un moyen permettant d'estimer ces mesures dans des conditions d'application du test sur le terrain. Ceci tient automatiquement compte des rendements réels du laboratoire incluant l'efficacité de son contrôle de qualité. Astudillo et Da Silva (1984) ont utilisé un raisonnement semblable pour l'étude d'un seuil d'interprétation pour la sérologie de la fièvre aphteuse.

La dynamique de la maladie et de la relation agent-hôte dans la population varie car une certaine adaptation se produit quand les souches de l'agent sont stables. Des changements peuvent se produire soudainement quand une nouvelle souche apparaît. Cette dynamique normale des populations de références saines et malades fait que la réaction immunitaire des animaux a tendance à réagir différemment au test avec le temps. Ceci a aussi pour conséquence de changer l'efficacité du test lors de la prise de décision diagnostique dans cette population. Il est probable que dans bien des cas il existe plus de deux sous-populations dans la grande population. Par exemple, il peut y avoir trois sous-populations, soit les naifs par rapport à la maladie, les cliniquement malades et les guéris qui gardent mémoire du contact avec la maladie. Dans ce cas, il est encore possible d'utiliser un modèle de distribution mixte constituée par un mélange de trois normales. Ceci donne une distribution à huit paramètres. Parker et al. (1990) ont rapporté un exemple semblable avec une sérologie appliquée au parvovirus B19. Leur but était d'identifier le seuil pour l'interprétation de la sérologie et le modèle s'est avéré utile pour solutionner cet important problème. En effet, une fois les moyennes et les variances des sous-populations connues il devient facile d'établir la relation entre seuil et sensibilité et spécificité.

Dans le cas d'une situation où il n'y a que deux sous-populations (sains et malades) dont les malades ont une tendance centrale plus élevée que les sains, la sensibilité est de un moins la probabilité d'observer un résultat du test plus haut que le seuil dans la distribution de fréquence des malades alors que la spécificité est la probabilité d'observer un résultat plus faible que le seuil dans la distribution de fréquence des sains. Ces valeurs peuvent être obtenues directement de tables statistiques de la distribution normale. L'application du seuil permet d'obtenir la probabilité d'être test positif et l'utilisation de la formule suivante nous donne un estimé de la prévalence. La valeur estimée des paramètres de la distribution mixte permet donc d'obtenir sensibilité, spécificité et prévalence en utilisant les conditions d'application du test sur le terrain.