Analysis of Diagnosability

Carlos Alberto de Braganca Pereira; Luis Raul Pericchi

Applied Statistics, Vol. 39, No. 2. (1990), pp. 189-204.

Stable URL:
http://links.jstor.org/sici?sici=0035-9254%281990%2939%3A2%3C189%3AAOD%3E2.0.CO%3B2-W

Applied Statistics is currently published by Royal Statistical Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/rss.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org
Sat Jan 26 20:50:32 2008

Appl. Statist. (1990)
39, *No.* 2, *pp.* 189–204

Analysis of Diagnosability

By CARLOS ALBERTO DE BRAGANÇA PEREIRA

Universidade de São Paulo, Brazil

and LUIS RAUL PERICCHI†

Universidad Simon Bolivar, Caracas, Venezuela

[Received June 1986. Revised May 1989]

SUMMARY

The diagnostic probabilities of having a disease based on possible responses (indicants) to a clinical question (tests, signs or symptoms) are generally given without reference to their precision. Here, a Bayesian approach is used to provide a full analysis of the diagnostic probabilities, the weights of evidence provided by each indicant and the average weight of evidence (diagnosability) provided by the question. The method is extended to a sequence of questions in which a particular response may influence whether a subsequent question is asked. The role of imprecise diagnostic probabilities in decision making is discussed.

Keywords: Bayes factor; Clinical indicant; Diagnosability; Dirichlet distribution; Divergence; Weight of evidence

1. Introduction

The quantities of interest in a medical diagnostic procedure are the probabilities of having or not having a disease, i.e. the *diagnostic probabilities*. These quantities may change their values in accordance with the diagnostic ability of the observed evidence. Evidence is produced by responses (called *indicants*) to clinical questions (tests, signs or symptoms). By *diagnosability*, a term introduced by Card (1967), we mean the expected ability of a question to modify the diagnostic probabilities. As an extension, it is said that one disease is more diagnosable than another if the set of questions relevant to the former provide on average a more drastic change in the physician's judgment (diagnostic probabilities) than the latter with its respective set of questions.

The indicant parameters related to a question are sensitivity and specificity which represent respectively the frequency of positive response among the patients having the disease and the frequency of negative response among the patients that do not have the disease. Vecchio (1966), using Bayes's formula and data from a simple contingency table, obtained the calibrated diagnostic probabilities. However, the variability of the estimates of the indicant parameters was not recognized since they were considered as the true values of these parameters. Consequently, the covariance structure of the quantities of interest, the diagnostic probabilities, was lost. Owing to

†*Address for correspondence*: Departamento de Matematica, Universidad Simon Bolivar, Apartado 89000, Caracas 1080 A, Venezuela.

© 1990 Royal Statistical Society 0035–9254/90/39189 $2.00

sample size restrictions, most of the published discussions on predictive values of responses ignore variability. Exceptions are Rigby (1982) and Critchley and Ford (1985), both in the context of multivariate normal populations.

Section 2 presents a fully Bayesian approach to Vecchio's set-up. The range of the diagnostic probabilities is completely described and their bivariate posterior density is obtained. This enhances knowledge of the covariance structure and all the information contained in each single indicant and in the expert prior information. The question example analysed in Section 2 is the following: the disease is cancer of the endometrius (whose states, presence and absence, are denoted by D and D' respectively) and the indicants are virginity, denoted by E^+, and its absence, denoted by E^-. The data (Table 1) were obtained by the Gynaecological Clinic of the Medical School of the São Paulo University in women of similar mature ages.

To measure diagnosability one is naturally led to the Bayes factor (BF), the ratio between the indicant parameters. The logarithm of the BF, as quite generally and cogently put forward by Good (1950, 1985), must be understood as the weight of evidence. Its evaluation gives the score of the evidence provided by the indicant. The expected weight of evidence (Good, 1950) is considered here as the diagnosability from a question or a set of questions. In Section 3 we show how to evaluate these quantities and their variabilities. The extension to more than one question, without assuming independence of questions (see de Dombal *et al.* (1974)), is discussed in Section 4. The analysis can be kept simple even for handling special kinds of nested missing data. In some cases questions have a natural order for response and depending on the indicant obtained for a specified question the subsequence may not be observed.

The data described by Table 2 and analysed in Section 4 are from the Hospital das Clínicas, São Paulo. They refer to 100 children with biliar obstruction which can take two forms: intrahepatic, D, and extrahepatic, D'. To help to discriminate between the states D and D', two clinical tests are available. Both tests have positive and negative outcomes. Thus evidence is one of the four possible indicants denoted by E^{++}, E^{+-}, E^{-+} and E^{--}, where the superscripts indicate the results of the two tests. Seven of the

TABLE 1
Incidence of virginity and cancer of the endometrius

State	E^+	E^-	*Sample size*
D	$x = 20$	$m - x = 130$	$m = 150$
D'	$n - y = 3$	$y = 147$	$n = 150$

TABLE 2
Frequency of indicants in children with biliar obstruction

State	E^{++}	E^{+-}	$E^{+\cdot}$	E^{-+}	E^{--}	$E^{-\cdot}$	*Sample size*
D	$x^{++} = 28$	$x^{+-} = 12$	$x^{+\cdot} = 40$	$x^{-+} = 1$	$x^{--} = 4$	$x^{-\cdot} = 5$	$m = 45$
D'	$y^{++} = 2$	$y^{+-} = 9$	$y^{+\cdot} = 11$	$y^{-+} = 2$	$y^{--} = 35$	$y^{-\cdot} = 37$	$n = 48$

children with state D took only the first test and all responded positively to it. Table 2 shows the results for the remaining 93 children. Among the children who took both tests, we indicate the frequency of children with positive (negative) response in the first test by $E^{+\cdot}$ ($E^{-\cdot}$) and in the second test by $E^{\cdot +}$ ($E^{\cdot -}$).

A major contribution on this subject is the work of Spiegelhalter and Knill-Jones (1984), where an impressively complicated problem was analysed in a simpler manner by putting together statistical and artificial intelligence techniques. The present article develops measures of assessment of evidential values of questions, and its variability, in computer-aided diagnosis and prognosis. The analysis remains quite simple, with minimal model assumptions, most useful for problems with a small number of questions, and a nested pattern of lack of responses (missing data).

The prior and posterior densities for the indicant parameters used throughout the paper are the Dirichlet densities and a particular mixture of them (Basu and Pereira, 1982). The justification of this choice is extensively discussed by Good (1965, 1967) and Novick and Grizzle (1965). We also consider the prior probability of having the disease $d = \Pr\{D\}$, as assessed by the physician since it represents his state of knowledge about his patient.

2. Joint Density for Diagnostic Probabilities

A patient consults a specialist who will investigate whether the patient has a disease, state D, or not, state D'. Before collecting further information, a prior probability, $d = \Pr\{D\}$, for the presence of the disease is assessed. This quantity translates, in some sense, the specialist's experience with other patients in similar circumstances. Looking for more information, the physician observes an indicant which is new evidence associated with the patient. Considering the first example of Section 1, about cancer of the endometrius, the parameters of interest (the diagnostic probabilities) are $\pi = \Pr\{D|E^+\}$ and $\theta = \Pr\{D'|E^-\}$, i.e. the predictive values of positive and negative responses respectively. However, the likelihood depends on d, $p = \Pr\{E^+|D\}$ and $q = \Pr\{E^-|D'\}$. d, p and q represent respectively the uncertainty of the clinician about the patient's state, the frequency of ill subjects who respond positively and the frequency of well subjects who respond negatively. Ill and well here are only related to the disease being studied and p and q are susceptible to selection effects (Dawid, 1976).

As a function of (p, q), the likelihood is proportional to

$$L^*(p, q) = p^x(1-p)^{m-x}q^y(1-q)^{n-y}. \tag{2.1}$$

To evaluate π and θ from d and (p, q), it is necessary to restrict ourselves to $p \neq 1-q$, i.e. the state of the patient D or D' and the evidence E^+ or E^- should be dependent; otherwise the indicants would be completely useless when searching to predict the state of the patient.

By Bayes's formula

$$\begin{aligned} \pi &= \frac{pd}{pd+(1-q)(1-d)}, \\ \theta &= \frac{q(1-d)}{q(1-d)+(1-p)d}, \end{aligned} \tag{2.2}$$

or, equivalently,

$$p = \frac{\pi\{\theta-(1-d)\}}{d(\pi+\theta-1)}, \qquad q = \frac{\theta(\pi-d)}{(1-d)(\pi+\theta-1)}. \tag{2.3}$$

Recall that $\pi \neq 1-\theta$ since this is equivalent to $p \neq 1-q$.

d is a convex combination of π and $1-\theta$ since $d = \pi \Pr\{E^+\} + (1-\theta)\Pr\{E^-\}$ and $\Pr\{E^-\} = 1 - \Pr\{E^+\}$. Consequently, (π, θ) must lie in the set $\Omega = \Omega_1 \cup \Omega_2$ where

$$\Omega_1 = \{(\pi, \theta);\ 0 \leqslant \pi \leqslant d \leqslant 1-\theta \leqslant 1 \text{ and } \pi \neq 1-\theta\},$$
$$\Omega_2 = \{(\pi, \theta);\ 0 \leqslant 1-\theta \leqslant d \leqslant \pi \leqslant 1 \text{ and } \pi \neq 1-\theta\}.$$

Fig. 1 illustrates the space Ω, the range of the quantities of interest.

Let $L(\pi, \theta)$ denote the likelihood function. It is clear that a good indicant will produce an L concentrated on Ω_2 (i.e. L may be considered nil in Ω_1). However, for large m and n, it is desirable that L concentrates itself in a region near (1, 1). If L is concentrated on Ω_2 (or alternatively on Ω_1), the ranges of π and θ may be very different depending on the value of d. For example, for d near zero, the set of possible values of θ is much smaller than that of π, whenever L is concentrated on Ω_2. In this case, estimates of θ might have smaller variances (although not necessarily) than estimates of π. Variances of estimates were not given by Vecchio (1966) who presented only point estimates for π and θ. Since the maximum likelihood estimate (MLE) is invariant, to obtain the MLE of (π, θ), it is sufficient to put

$$(\hat{p}, \hat{q}) = \left(\frac{x}{m}, \frac{y}{n}\right)$$

in equations (2.2).

For our example, taking $d = 0.1$,

$$(\hat{p}, \hat{q}) = \left(\frac{2}{15}, \frac{49}{50}\right)$$

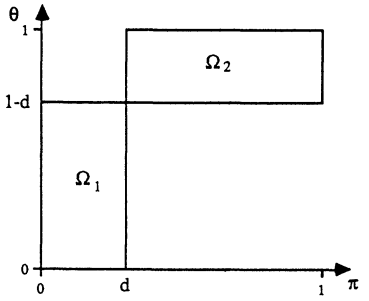

Fig. 1. Range of diagnostic probabilities

and

$$(\hat{\pi},\ \hat{\theta}) = (0.426,\ 0.911).$$

Note that the probability of D changes from 0.1 to 0.426 if E^+ is obtained. In contrast, the probability of D' changes from 0.9 to 0.911 if E^- is obtained. This suggests that, although $\hat{\theta}$ may be more precise than $\hat{\pi}$, $\hat{\pi}$ can be more informative than $\hat{\theta}$ (owing to the amount of change produced in the probabilities of D and D').

Proceeding with a Bayesian analysis, as a conjugate prior distribution for (p, q), we consider p and q independently distributed as beta distributions with parameters (a, b) and (a', b') respectively. Consequently, *a posteriori*, p and q are independently distributed in accord with beta distributions with parameters $(A, B) = (a+x, b+m-x)$ and $(A', B') = (a'+y, b'+n-y)$ respectively.

The Jacobian of the transformation from (p, q) to (π, θ) is given by

$$\|J\| = \frac{(\pi-d)\{\theta-(1-d)\}}{d(1-d)}\ |\pi+\theta-1|^{-3},$$

where $|x|$ is the absolute value of x. The posterior joint density of (π, θ) is then

$$f(\pi,\theta) = \frac{1}{B(A,B)\,B(A',B')}\ \frac{\pi^{A-1}(1-\theta)^{B-1}\theta^{A'-1}(1-\pi)^{B'-1}}{d^{A+B-1}(1-d)^{A'+B'-1}}$$
$$\times \frac{|\theta+d-1|^{A+B'-1}\,|\pi-d|^{B+A'-1}}{|\pi+\theta-1|^{A+A'+B+B'-1}}$$

for $(\pi, \theta) \in \Omega = \Omega_1 \cup \Omega_2$ and zero otherwise, where $B(\cdot, \cdot)$ denotes the beta function.

Consider now the data of Table 1 and a uniform prior for (p, q), i.e. in the prior $a = b = a' = b' = 1$, and in the posterior $A = 21$, $B = 131$, $A' = 148$ and $B' = 4$. The posterior mode with these parameters is $(\tilde{\pi}, \tilde{\theta}) = (0.321, 0.909)$. Note the sensitivity with respect to the prior when comparing the values of $\hat{\pi}$ and $\tilde{\pi}$. The first may be viewed as the posterior mode in the case of a uniform prior for (π, θ) and the second as the posterior mode when the uniform prior is for (p, q). Fig. 2 displays the mode and five credible sets (99%, 95%, 90%, 80% and 50% credibility).

Figs 3 and 4 are the posterior marginal densities of π and θ respectively. They show the difference of variability between these two quantities. The modes of these densities are respectively 0.35 and 0.91. The two corresponding 95% credible sets are (0.17, 0.67) and (0.905, 0.916). These results were obtained by numerical integration.

The algebraic difficulties that we must face do not allow us to give the analytical expression of the marginal densities of π and θ. This is due mainly to the uncommon range Ω of (π, θ). Even the moments of π and θ are not easy to obtain directly. However, as shown in what follows some alternative quantities of interest have explicit analytical expressions for their estimators.

3. Measuring Evidence

3.1. *Weight of Evidence*

The posterior odds in favour of the states D and D' are respectively

$$\varnothing_+ = \varnothing(D|E^+) = \frac{\pi}{1-\pi}$$

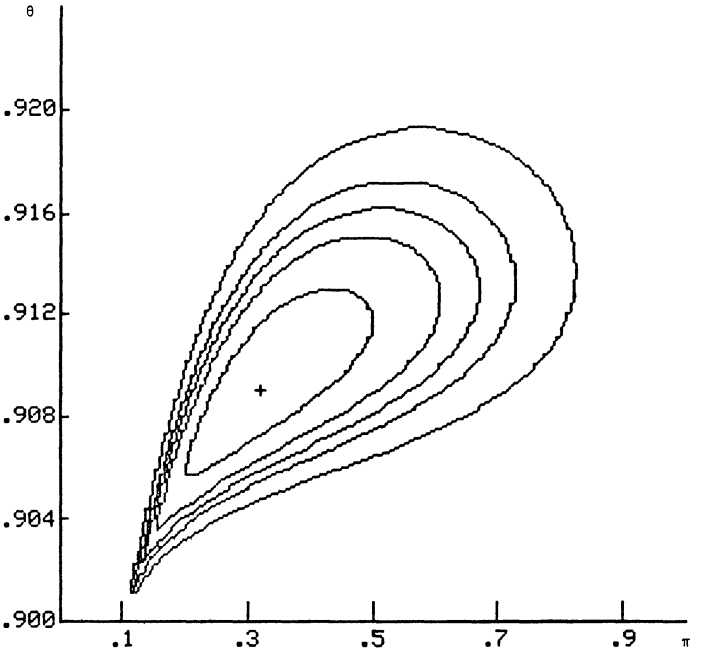

Fig. 2. Credible regions for (π, θ): 50%, 80%, 90%, 95% and 99% credibility

and

$$\varnothing'_- = \varnothing'_-(D' \mid E^-) = \frac{\theta}{1-\theta}.$$

The corresponding prior odds are $\varnothing = d/(1-d)$ and $\varnothing' = 1/\varnothing$.

It has already been postulated that to evaluate the diagnostic ability of the evidence, E^+ or E^-, we should focus on the change from $\varnothing$ to $\varnothing_+$ and from $\varnothing'$ to $\varnothing'_-$. This is linked with the weight of evidence provided by E^+ (E^-) in favour of D (D'), denoted by $\omega = \omega(D; E^+)$ $(\omega' = \omega(D'; E^-))$. Good (1968), p. 141, proved that the weight of evidence, to obey reasonable requirements, ought to be an increasing function of the BF, the ratio of posterior to prior odds, i.e. ω and ω' are respectively increasing functions of $r = \varnothing_+/\varnothing = p/(1-q)$ and $r' = \varnothing'_-/\varnothing' = q/(1-p)$.

The BFs r and r' can be seen as likelihood ratios. The relationships between the BFs with sensitivity p and specificity q is quite simple and intuitively appealing. The usual cross-product, useful in testing independence and measuring association, is simply

$$\rho = \frac{pq}{(1-p)(1-q)} = rr' = \frac{\pi\theta}{(1-\pi)(1-\theta)}.$$

Since p and q are considered as independent beta variables, the estimation of r, r' and ρ is not difficult. The posterior expected values, taken as the Bayes estimator for these quantities, together with their variances are given in result 7 in Appendix A.

Distributional results which permit the use of numerical methods to obtain credible intervals can be found in Aitchison and Bacon-Shone (1981), Altham (1969) and Weisberg (1972).

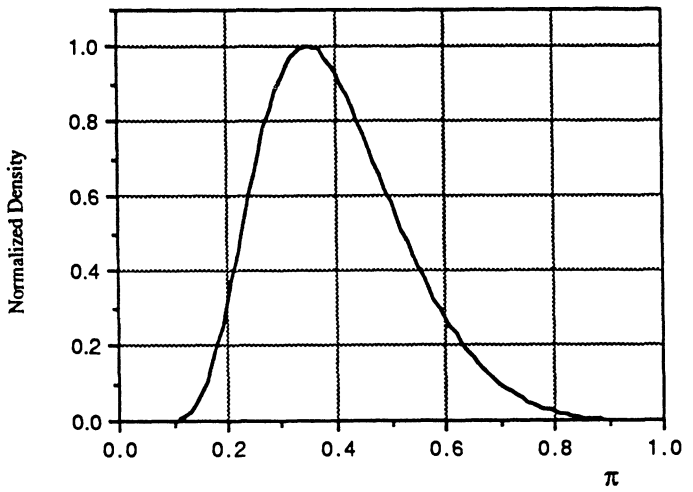

Fig. 3. Posterior marginal density for π normalized by the mode

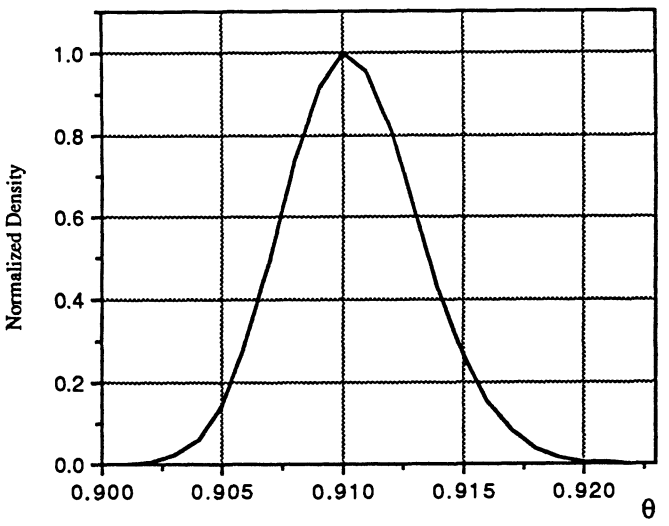

Fig. 4. Posterior marginal density for θ normalized by the mode

As a consequence of the natural requirement of additivity of information, Good (1968) proves that $\omega(\omega')$ is the logarithm, the base depending on the unit of measurement chosen, of $r(r')$, i.e. the weight of evidence is the logarithm of the BF. For simplicity, we use only natural logarithms in this paper. Good (1950) also points out that the expected value of the weight of evidence is more meaningful than that of the BF. The possible weights of evidence are

(a) in favour of D, $\omega(D; E^+) = \omega = \ln r$ and $\omega(D; E^-) = -\omega' = \ln(1/r')$ and
(b) in favour of D', $\omega(D'; E^+) = -\omega = \ln(1/r)$ and $\omega(D'; E^-) = \omega' = \ln r'$.

Using result 5 (Appendix A) we have the following results, where $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and trigamma functions respectively.

Result 1. With the distributions considered we have

(a) $\bar{\omega} = E(\omega | x, y) = \psi(A) - \psi(A+B) - \{\psi(B') - \psi(A'+B')\}$ and
(b) $\text{var}(\omega | x, y) = \psi'(A) - \psi'(A+B) + \psi'(B') - \psi'(A'+B')$.

To obtain $\bar{\omega}' = E(\omega' | x, y)$ and $\text{var}(\omega' | x, y)$ it is sufficient to interchange A with A' and B and B' throughout.

If normal approximations for the distributions of ω and ω' are appropriate, credible sets are easily obtained. The reader should compare the exact expressions (a) and (b), assuming $a = b = a' = b' = 1$, with the approximate expressions of Spiegelhalter and Knill-Jones (1984), pp. 46 and 76. Of course the expert is not restricted to the choice of uniform priors. Our only restriction here is the use of independent beta priors for p and q. The method described here can be used with other classes of priors. However, closed simple expressions such as those given here may not be obtained. The influence of the prior choice may be evaluated by using the divergence measure defined in the following.

Regarding the estimation of the log-odds, represent the prior log-odds by $\omega_0 = \ln\{d/(1-d)\}$. Hence, $E(\ln \varnothing_+ | x, y) = \omega_0 + \bar{\omega}$ and $E(\ln \varnothing_- | x, y) = -\omega_0 + \bar{\omega}'$. The variances of these log-odds are $\text{var}(\omega | x, y)$ and $\text{var}(\omega' | x, y)$ since ω_0 is fixed.

We next define a measure of the ability of a question to discriminate in favour of $D\,(D')$, given that $D\,(D')$ is the true state. This measure is the conditional expectation of $\omega\,(\omega')$ given p, q and the state of the patient, D or D'. For the two possible cases D and D', these conditional expectations are denoted by α and α' respectively, where $\alpha = p\omega - (1-p)\omega'$ and $\alpha' = q\omega' - (1-q)\omega$. The divergence measure (Jeffreys, 1946; Kullback, 1959) is defined as

$$\delta = \alpha + \alpha' = (p+q-1)\ln\rho.$$

This measure is non-negative, symmetric with respect to D and D', and is interpreted as the mean plausibility gained about the correct state of nature. Also $\delta = 0$ if and only if $\rho = 1$ which is the case of independence between state of the patient (D or D') and response (E^+ or E^-).

Finally, the overall diagnostic ability or *diagnosability* of a question is defined as

$$\mathscr{D} = E(\delta | x, y),$$

the posterior expected value of the divergence.

To present the next result we recall the following expected values

$$E(p | x, y) = \bar{p} = \frac{A}{A+B},$$

$$E(q | x, y) = \bar{q} = \frac{A'}{A'+B'}$$

and

$$E(\ln\rho | x, y) = \bar{\omega} + \bar{\omega}' = \psi(A) + \psi(A') - \psi(B) - \psi(B').$$

Result 2. With the distributions considered we have

(a) $$\bar{\alpha} = E(\alpha | x, y) = \bar{p}\bar{\omega} - (1 - \bar{p})\bar{\omega}' + \frac{1}{A + B}$$

and

(b) $$\mathscr{D} = (\bar{p} + \bar{q} - 1)\, E(\ln \rho | x, y) + \frac{1}{A + B} + \frac{1}{A' + B'}.$$

To obtain $\bar{\alpha}'$ from (a) it is sufficient to interchange A with A' and B with B', and to substitute $\bar{q}$ for $\bar{p}$, $\bar{\omega}$ for $\bar{\omega}'$ and $\bar{\omega}'$ for $\bar{\omega}$.

The proof of this result follows from result 5 (Appendix A). The variances of ω, ω' and δ may also be exactly obtained by using results 5 and 6.

Let Q_1 and Q_2 represent two different questions with diagnosabilities $\mathscr{D}_1$ and $\mathscr{D}_2$ respectively, satisfying $\mathscr{D}_1 < \mathscr{D}_2$. We say that on average Q_2 has more diagnostic ability than Q_1 has. An application is given for the example in Section 4.2.

Next, the example of Table 1 is reviewed under the light of the results of this section.

3.2. *Example*

Returning to the cancer of the endometrius example, we have assumed in Section 2 that $d = 0.1$ and a uniform distribution ($a = b = a' = b' = 1$) in the unit square as the referential prior for (p, q). Denoting the standard deviation by $\sigma(\cdot)$ and noting that $\omega_0 = \ln d - \ln(1 - d) = -2.2$ we obtain the following results:

(a) related to D, $\bar{\omega} = 1.76$, $\omega_0 + \bar{\omega} = -0.44$, $\sigma(\omega) = 0.565\,13$ and $\bar{\alpha} = 0.1448$;
(b) related to D', $\bar{\omega}' = 0.12$, $-\omega_0 + \bar{\omega}' = 2.32$, $\sigma(\omega') = 0.035\,21$ and $\bar{\alpha}' = 0.0794$.

Note that $\bar{\omega}$ is about 15 times $\bar{\omega}'$ and $\bar{\alpha}$ is almost twice $\bar{\alpha}'$. Since $\bar{\omega}$ and $\bar{\alpha}$ are related to the ability to detect D and $\bar{\omega}'$ and $\bar{\alpha}'$ to D', we suggest that if D is true more will be learned with the response about virginity than if D' is true. However, since $\sigma(\bar{\omega})$ is about 16 times $\sigma(\bar{\omega}')$, the precision of the estimates related to D' is higher than the precision of those related to D. All these results are in complete agreement with the direct and complicated analysis presented in Section 2. The positive evidence E^+ seems to be much more informative than the negative evidence E^-, despite the fact that $\bar{p} = 0.1378 < \bar{q} = 0.974$. However, the negative evidence produces more precise results than the positive evidence.

The analysis described in this section is exact, flexible and simple. In what follows we introduce an extension for the case where two or more questions are involved.

4. More than One Source of Evidence

4.1. *The Model*

Most of the recent work on this subject assumes implicitly or explicitly that, depending on the true state of the patient, different questions produce independent evidence (e.g. Vecchio (1966) and de Dombal *et al.* (1974)). This assumption greatly simplifies the analysis, but when it is not a fair assumption the resulting accumulated weight of evidence may be completely misleading. This section describes a model that

is free of the independence assumption yet retains much of the simplicity of the previous analyses.

For simplicity and clarity we only develop the model for two questions since the general case is a simple extension.

Suppose that after having responded to a clinical question patients are asked to respond to a second question. The four possible pieces of evidence in this case are E^{++}, E^{+-}, E^{-+} and E^{--}. The first (second) superscript indicates the response, positive or negative, to the first (second) question. For the patients who responded to both questions, $E^{+\cdot}$ and $E^{-\cdot}$ ($E^{\cdot+}$ and $E^{\cdot-}$) indicate their response to only the first (second) question. Consider that, owing to cost or difficulty in obtaining a response for the second question, some of the patients do not respond to the second question. (The case of full response is a particular case of this non-response data model.) Fig. 5 introduces the notation for the set of patients who responded to both questions.

The overall likelihood function for the complete data is

$$\{p^x(1-p)^{m-x}q^y(1-q)^{n-y}\}\ \{p_+^{x^{++}}(1-p_+)^{x^{+-}}q_+^{y^{++}}(1-q_+)^{y^{+-}}\}$$
$$\times\ \{p_-^{x^{-+}}(1-p_-)^{x^{--}}q_-^{y^{-+}}(1-q_-)^{y^{--}}\}$$

where, as before, m (n) is the total number of patients in D (D') who responded to the first question and x (y) is the number of patients in D (D') who responded positively (negatively) to the first question. The first factor involves all patients in the sample whereas the other two factors involve only patients who responded to both questions.

To proceed with the Bayesian analysis, we need to specify an *a priori* joint distribution for the parameters of Fig. 5. The candidate that generalizes our previous choice is given by two independent Dirichlet distributions for $(p_{++}, p_{+-}, p_{-+}, p_{--})$ and $(q_{++}, q_{+-}, q_{-+}, q_{--})$ whose respective parameters are represented by (a_+, a_-, b_+, b_-) and (a'_+, a'_-, b'_+, b'_-).

This prior choice is equivalent to considering the quantities p, q, p_+, q_+, p_- and q_- as independent beta variables with parameters $(a, b) = (a_+ + a_-, b_+ + b_-)$, $(a', b') = (a'_+ + a'_-, b'_+ + b'_-)$, (a_+, a_-), (b'_-, b'_+), (b_+, b_-) and (a'_-, a'_+) respectively (see Basu and Pereira (1982)).

A posteriori, the quantities p, q, p_+, q_+, p_- and q_- are independent beta variables with parameters $(A, B) = (A_+ + A_-, B_+ + B_-)$, $(A', B') = (A'_+ + A'_-, B'_+ + B'_-)$, (A_+, A_-), (B'_-, B'_+), (B_+, B_-) and (A'_-, A'_+) respectively, where

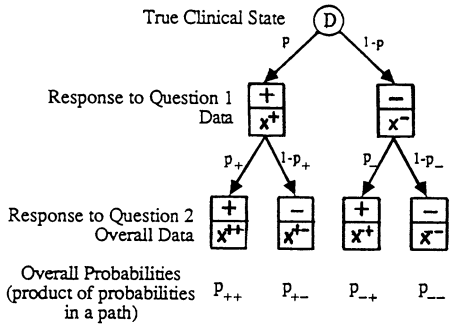

Fig. 5. Description of the data for the case of two questions (for D', replace $1-p$ by q, p by $1-q$ and x by y)

$$A = a + x, \qquad B = b + m - x, \qquad A' = a' + y, \qquad B' = b' + n - y,$$

$$A_+ = a_+ + x^{+\,+}, \qquad A_- = a_- + x^{+\,-}, \qquad B_+ = b_+ + x^{-\,+}, \qquad B_- = b_- + x^{-\,-},$$

$$A'_+ = a'_+ + y^{-\,-}, \qquad A'_- = a'_- + y^{-\,+}, \qquad B'_+ = b'_+ + y^{+\,+}, \qquad B'_- = b'_- + y^{+\,-}.$$

Since A, B, A' and B' are not necessarily equal to $A_+ + A_-$, $B_+ + B_-$, $A'_+ + A'_-$ and $B'_+ + B'_-$ respectively, we conclude that, *a posteriori*, the original parameters are not distributed as independent Dirichlet distributions but as independent mixtures of Dirichlet distributions (Basu and Pereira, 1982).

With the representation of the posterior distribution as stated, the computation of the weights of evidence, their estimates and their variances, remains fairly simple. The weights of evidence are introduced by Fig. 6. The overall weight of evidence is the sum of the weight of evidence of the response for the first question plus the conditional weight of the evidence of the response for the second question given the response for the first. For example, $\omega_{+\,+} = \omega + \omega_+$, where $\omega_+ = \ln\{p_+/(1 - q_+)\}$.

The estimate of the overall weight of evidence is equal to the sum of the estimate of the first factor, given in result 1(a), plus the estimate of the conditional weight of evidence which is given by the following property that follows from result 5 (Appendix A).

Result 3. With the distributions considered we have

(a) $\bar{\omega}_+ = \psi(A_+) - \psi(A_+ + A_-) - \{\psi(B'_+) - \psi(B'_+ + B'_-)\}$ and
(b) $\bar{\omega}'_+ = \psi(B'_-) - \psi(B'_+ + B'_-) - \{\psi(A_-) - \psi(A_+ + A_-)\}$.

To obtain $\bar{\omega}_-$ and $\bar{\omega}'_-$ from (a) and (b) respectively, it is sufficient to interchange A and B throughout. To obtain the variances it is sufficient to replace ψ and ψ' and the operation $-$ with $+$ before the brackets.

Another feature of the parameterization considered here is that it produces simple estimates of BFs. For example, if both responses are positive, then the BF is

$$rr_+ = \frac{p}{1-q}\,\frac{p_+}{1-q_+}.$$

The two ratios are independent and the expectation of the first is given by result 7(a). Since

$$E(r_+ \,|\, x, y) = \frac{(B' - 1)A_+}{(B'_+ - 1)A},$$

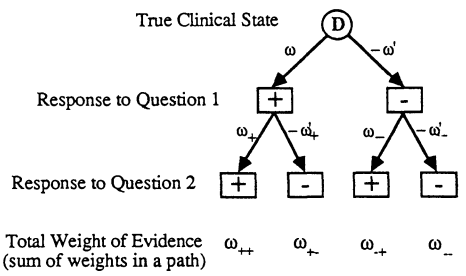

Fig. 6. Weights of evidence in the case of two questions

the estimate of the BF is

$$E(rr_+ \mid x, y) = \frac{(A' + B' - 1)A_+}{(B'_+ - 1)(A + B)}.$$

To obtain the extended version of α, α' and δ, denoted respectively by $\mathscr{A}$, $\mathscr{A}'$ and Δ, we follow Good (1983). Related to the ability of the sensitivity and the specificity we have, respectively, the following expressions:

$$\mathscr{A} = \alpha + p\alpha_+ + (1-p)\alpha_- \tag{4.1}$$

and

$$\mathscr{A}' = \alpha' + (1-q)\alpha'_+ + q\alpha'_-, \tag{4.2}$$

where α and α' are defined in Section 3.1 and $\alpha_+ = p_+\omega_+ - (1 - p_+)\omega'_+$, $\alpha_- = p_-\omega_- - (1-p_-)\omega'_-$, $\alpha'_+ = q_+\omega'_+ - (1-q_+)\omega_+$ and $\alpha'_- = q_-\omega'_- - (1-q_-)\omega_+$.

As before, the overall divergence is the sum of these two measures, i.e. $\Delta = \mathscr{A} + \mathscr{A}'$. The overall diagnosability is $\mathscr{D} = E(\Delta)$.

To estimate all these quantities we have the following result.

Result 4. With the distributions considered, we have

(a) $\overline{\mathscr{A}} = E(\mathscr{A}) = \bar{\alpha} + \bar{p}\{\bar{p}_+\bar{\omega}_+ - (1-\bar{p}_+)\bar{\omega}'_+ + (A_+ + A_-)^{-1}\} + (1-\bar{p})\{\bar{p}_-\bar{\omega}_- - (1-\bar{p}_-)\bar{\omega}'_- + (B_+ + B_-)^{-1}\}$,

(b) $\overline{\mathscr{A}'} = E(\mathscr{A}') = \bar{\alpha}' + (1-\bar{q})\{\bar{q}_+\bar{\omega}'_+ - (1-\bar{q}_+)\bar{\omega}_+ + (B'_- + B'_+)^{-1}\} + \bar{q}\{\bar{q}_-\bar{\omega}'_- - (1-\bar{q}_-)\bar{\omega}_- + (A'_- + A'_+)^{-1}\}$,

where $\bar{p}_+ = A_+/(A_+ + A_-)$, $\bar{p}_- = B_+/(B_+ + B_-)$, $\bar{q}_+ = B'_-/(B'_- + B'_+)$ and $\bar{q}_- = A'_-/(A'_- + A'_+)$.

Finally we have the overall measure of diagnosability given by

$$\mathscr{D} = E(\Delta) = \overline{\mathscr{A}} + \overline{\mathscr{A}'}.$$

To prove this property and to obtain the variances of $\mathscr{A}$, $\mathscr{A}'$ and Δ it is sufficient to use results 5 and 6 and to consider the fact that p, q, p_+, q_+, p_- and q_- are mutually independent. The expressions of the variances, although rather lengthy, are computationally straightforward.

The extension of the methodology described in this section for more than two questions, say $Q_1, Q_2, \ldots, Q_k$, follows the same lines if we have nested samples as in our case. For example, consider three questions Q_1, Q_2 and Q_3. Suppose also that the set of units that responds to Q_1 contains the set of units that also responds to Q_2 and that this last set contains the set of units that responds to Q_3 in addition to Q_2 and Q_3. The chain of independent quantities may be intuitively represented by

$$\text{Hierarchical chain} \begin{cases} D \to p \begin{cases} p_+ \begin{cases} p_{++} \\ p_{+-} \end{cases} \\ p_- \begin{cases} p_{-+} \\ p_{--} \end{cases} \end{cases} \\ D' \to q \begin{cases} q_+ \begin{cases} q_{++} \\ q_{+-} \end{cases} \\ q_- \begin{cases} q_{-+} \\ q_{--}. \end{cases} \end{cases} \end{cases}$$

The nested data handled by the methodology developed is a particular case of missing data. For a more general kind of missing observations we can follow the same approach but must use complicated techniques as described in Dickey *et al.* (1987).

As a final remark, the backward elimination procedure in a logistic model, as performed by Spiegelhalter and Knill-Jones (1984) to counterbalance the independence assumption, is unnecessary for the present method. These authors agreed with the argument of Copas, in the discussion of that paper, that the elimination procedure is unsatisfactory for deriving a predictive model.

4.2. *Example: Two Questions*

Consider the two-questions example of Section 1. From Table 2 and the fact that seven patients have responded only to the first test, we have the following results:

$$x = 47; \qquad m = 52; \qquad y = 37; \qquad n = 48.$$

For a referential analysis we assume that $a_+ = b_+ = a_- = b_- = a'_+ = b'_+ = a'_- = b'_- = \frac{1}{2}$. Since in this case $a = b = a' = b' = 1$, this is consistent with the prior specification in the example of Section 3.2, namely uniform priors for the marginal parameters. The medical doctors involved with the problem agreed with the statement that $\Pr(D) = 0.6$, which implies that $\omega_0 = 0.405$.

The final results based on these prior considerations are included in Table 3. Many of the standard methods assume conditional independence between the responses given the state of nature, D or D'. For comparison, in parentheses below the results, we present the relative change as a percentage if the so-called 'independent Bayes' approach is used.

By analysing the results in Table 3, since $\overline{\mathscr{A}}$ and $\overline{\mathscr{A}}'$ are very close, the abilities of sensitivity and specificity are about the same. $\sigma(\omega_{-+})$ deserves particular attention as it is almost twice $|\omega_{-+}|$ and is the highest among the standard deviations. This is in accord with the fact that E^{-+} contains only three observations and that this cell shows the greatest discrepancy with the independent Bayes approach. If conditional independence is used we would have as Bayes estimates $\tilde{\omega}_{-+} = -0.06$ and as the posterior standard deviation $\tilde{\sigma}(\omega_{-+}) = 0.62$. This discussion shows strong disagreement between the two approaches. In effect, the independent Bayes approach underestimates the dispersion of all cells and overestimates the sensitivity and the specificity

TABLE 3
Numerical results for the problem of biliar obstruction†

Posterior	*Weight of evidence*				*Weight of evidence in average*	
	ω_{++}	ω_{+-}	ω_{-+}	ω_{--}	$\mathscr{A}$	$\mathscr{A}'$
Mean (estimate)	2.711 (21%)	0.369 (11%)	−0.744 (−93%)	−2.244 (30%)	1.626 (18%)	1.485 (26%)
Standard deviation	0.695 (−24%)	0.390 (−15%)	1.177 (−47%)	0.488 (−6%)	— —	— —

† The changes by using the independent Bayes approach are given in parentheses.

abilities of the joint effect of the two facets. The independent Bayes approach unduly claims a diagnosability (3.7898) which is 22% higher than the more realistic $\mathscr{D} = 3.111$ claimed by our model.

In conclusion, our analysis which leads to Table 3 is exact, giving the measures of the evidence and their dispersion; it is simple and easy to interpret. Furthermore, it is free from the 'conditional independence' assumption, which might well be highly misleading, as it is in the present example.

4.3. *Final Observations*

As Good (1985) points out, the weight of evidence is not a utility function but rather a 'quasi-utility'. It does not state the cost of different errors. However, ω is crucial on its own merits by weighing the plausibility gained by each state of nature and by rating patients or questions. Sometimes it is satisfactory to base a decision on the probabilities alone because in such cases the consequences of errors of misclassification are roughly equal. An instance of this is the following. In practice, owing to restrictions of time, budget etc., we might have to choose between different tests. The data from Table 2 give us the opportunity to compare the diagnostic ability of the first test ($\mathscr{D}_1 = 2.05$) with that of the second test ($\mathscr{D}_2 = 1.586$). If the cost of misclassification is about the same, then the first test should be preferred to the second. As a second instance, Spiegelhalter and Knill-Jones (1984) classify the patients using only the weight of evidence. In that work, however, the threshold to decide whether there is enough evidence of illness for any patient depends only on the value of the conditional probability of D given the evidence and it is the same for all patients. A more reasonable approach appears to be the inclusion of the value of the variance of ω in the definition of the threshold. Hence, this alternative approach would consider a specific threshold for each patient since evidence varies with patients. Tentatively we suggest the following rule: taking into consideration losses and accurateness of prediction, define adequately constants t, t', k and k' and then

(a) if $\bar{\omega} \geqslant t + k\sigma(\omega)$ and $\bar{\omega}' < t' + k'\sigma(\omega')$, classify the patient in D,
(b) if $\bar{\omega} < t + k\sigma(\omega)$ and $\bar{\omega}' \geqslant t' + k'\sigma(\omega')$, classify the patient in D' and
(c) if neither of the cases (a) and (b) is satisfied, new evidence must be sought, i.e. classify the patient in D or D' only if none or little conflicting evidence exists.

A limitation shared by other methods is the binary restriction for the states of nature, D and D'. Considering two possible diseases, the major difficulty is to decide whether the patient is free of both of them. This was recognized by Spiegelhalter and Knill-Jones (1984). The main limitation is to state a workable extension of the weight of evidence. Consider just one question with two different diseases, say D_1 and D_2, plus the absence of both, D'. Denoting $D = D_1 \cup D_2$, a convenient route is to consider first the BF of D versus D' and then the conditional BF of D_1 versus D_2 given D if D is eventually more plausible than D'. A study of such a route is left for a future work.

Acknowledgements

The financial support of three Brazilian research agencies, FAPESP, CNPq and FINEP, made possible the visit of the second author to the University of São Paulo and the visit of the first author to Richard Barlow at the University of California,

Berkeley. While completing this work, the authors were partially supported, respectively, by ARO (grant DAAG29-85-k-0208) at Berkeley and a Science and Engineering Research Council fellowship at Imperial College of Science and Technology, London. The referee's comments on early versions have greatly improved the paper. Marcio de Carvalho, our computer expert friend, kindly corrected our numerical results from an earlier version of the paper and wrote the new computer programs for the results and graphics of Section 2. The staff of the Informática Médica of the Hospital das Clínicas de São Paulo kindly presented the data of Section 4.2 to us. Dr A. F. M. Mitchell made several suggestions.

Appendix A

We list some basic standard theoretical results used throughout the paper. Consider two independent beta random variables p and q, with positive parameters (a, b) and (a', b') respectively. Denote by $\psi(\cdot)$ and $\psi'(\cdot)$ the digamma and the trigamma functions.

Result 5. If $\ln p$ represents the natural logarithm of p, then $E(\ln p) = \psi(a) - \psi(a+b)$ and $\text{var}(\ln p) = \psi'(a) - \psi'(a+b)$, where $\text{var}(\cdot)$ denotes variance. If q replaces p, then a' and b' will replace a and b respectively. If $1-p$ replaces p, then a and b will interchange (and analogously for $1-q$ replacing q).

Result 6.

$$E\{\ln p \ln(1-p)\} = \psi(a+b)\{\psi(a+b) - \psi(a) - \psi(b)\} + \psi(a)\,\psi(b) - \psi'(a+b).$$

If q replaces p, then a' and b' will replace a and b respectively.

Result 7. Let r, r' and ρ be as in Section 3 and A, B, A' and B' as in Section 2. With the distributions considered we have

(a) $$E(r|x,y) = \frac{A}{A+B}\,\frac{A'+B'-1}{B'-1} \quad \text{if } B' > 1,$$

(b) $$E\left(\frac{1}{r}\middle| x, y\right) = \frac{B'}{A'+B'}\,\frac{A+B-1}{A-1} \quad \text{if } A > 1,$$

(c) $$E(\rho|x,y) = \frac{A}{B-1}\,\frac{A'}{B'-1} \quad \text{if } B > 1 \text{ and } B' > 1,$$

(d) $$\text{var}(r|x,y) = E(r|x,y)\left\{\frac{A+1}{A+B}\,\frac{A'+B'-2}{B'-2} - E(r|x,y)\right\} \quad \text{if } B' > 2,$$

(e) $$\text{var}\left(\frac{1}{r}\middle| x, y\right) = E\left(\frac{1}{r}\middle| x, y\right)\left\{\frac{B'+1}{A'+B'}\,\frac{A+B-2}{A-2} - E\left(\frac{1}{r}\middle| x, y\right)\right\}$$
if $A > 2$ and

(f) $$\text{var}(\rho|x,y) = E(\rho|x,y)\left\{\frac{A+1}{B-1}\,\frac{A'+1}{B'-1} - E(\rho|x,y)\right\} \quad \text{if } B > 1 \text{ and } B' > 1.$$

To obtain the posterior expectations of r' and $1/r'$ from (a) and (b) respectively and their variances from (d) and (e) respectively, it is sufficient to interchange A with A' and B with B' throughout.

References

Aitchison, J. and Bacon-Shone, J. (1981) Bayesian risk ratio analysis. *Am. Statistn*, **35**, 254–257.

Altham, P. M. E. (1969) Exact Bayesian analysis of a 2×2 contingency table and Fisher's "exact" significance test. *J. R. Statist. Soc.* B, **31**, 261–269.

Basu, D. and Pereira, C. A. de B. (1982) On the Bayesian analysis of categorical data: the problem of non-response. *J. Statist. Planng Inf.*, **6**, 345–362.

Card, W. I. (1967) Towards a calculus of medicine. *Med. A.*, **9**, 9–21.

Critchley, F. and Ford, I. (1985) Interval estimation in discrimination: the multivariate normal equal covariance case. *Biometrika*, **72**, 109–161.

Dawid, A. P. (1977) Properties of diagnostic data distributions. *Biometrics*, **32**, 647–658.

Dickey, J. M., Jiang, J. M. and Kadane, J. B. (1987) Bayesian methods for censored categorical data. *J. Am. Statist. Ass.*, **82**, 773–781.

de Dombal, F. T., Leaper, D. J., Horrocks, J. C., Staniland, J. R. and McCann, A. P. (1974) Human and computer-aided diagnosis of abdominal pain: further report with emphasis on performance of clinicians. *Br. Med. J.*, **1**, 376–380.

Good, I. J. (1950) *Probability and the Weighing of Evidence*. London: Griffin.

——— (1965) *The Estimation of Probabilities*. Cambridge: Massachusetts Institute of Technology Press.

——— (1967) A Bayesian significance test for multinomial distributions (with discussion). *J. R. Statist. Soc.* B, **29**, 399–431.

——— (1968) Corroboration, explanation, involving probability, simplicity and a sharpened razor. *Br. J. Phil. Sci.*, **19**, 123–143.

——— (1983) *Good Thinking: the Foundations of Probability and its Applications*. Minneapolis: University of Minnesota Press.

——— (1985) Weight of evidence: a brief survey. In *Bayesian Statistics 2* (eds J. M. Bernado, H. DeGroot, D. V. Lindley and A. F. M. Smith), pp. 249–270. Amsterdam: North-Holland.

Jeffreys, H. (1946) An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond.* A, **186**, 436–461.

Kullback, S. (1959) *Information Theory and Statistics*. New York: Wiley.

Novick, M. R. and Grizzle, J. E. (1965) A Bayesian approach to the analysis of data from clinical trials. *J. Am. Statist. Ass.*, **60**, 81–96.

Rigby, R. A. (1982) A credibility interval for the probability that a new observation belongs to one of two multivariate normal populations. *J. R. Statist. Soc.* B, **44**, 212–220.

Spiegelhalter, D. J. and Knill-Jones, R. P. (1984) Statistical and knowledge-based approaches to clinical decision-support systems, with an application in gastroenterology (with discussion). *J. R. Statist. Soc.* A, **147**, 35–77.

Vecchio, T. J. (1966) Predictive value of a single diagnostic test in unselected populations. *New Engl. J. Med.*, **274**, 1171–1173.

Weisberg, H. I. (1972) Bayesian comparison of two ordered multinomial populations. *Biometrics*, **28**, 859–867.