CHAPTER 15

Influence Diagrams and Medical Diagnosis

Carlos A. de B. Pereira, *Universidade de São Paulo, Brazil*

ABSTRACT

Influence diagrams operations are used to solve the following problem:

A patient consults with a specialist who is going to start a search to discover whether the patient has a disease, D, or its absence, D'. Before collecting any further information, a prior probability, $d = \Pr\{D\}$, for the presence of the disease is assessed. Looking for more information, the physician observes an indicant (E = positive response or E' = negative response), which is a new evidence associated with the patient. The experience of the physician is in part represented by the data (x, y), where $x(y)$ is the number of positive (negative), respondents among all former patients having $D(D')$. The objective is to evaluate, for the new patient, the conditional probability of D (D') given that the patient responded positively (negatively) and also that the data (x, y) have been observed. Note that the likelihood depends upon the sensitivity, $\pi = \Pr\{E|D\}$, and the specificity, $\theta = \Pr\{E'|D'\}$. However the parameters of interest, the diagnostic probability, are $p = \Pr\{D|E\}$ and $q = \Pr\{D'|E'\}$. In another context the same problem is discussed by Pereira and Pericchi (1990).

15.1 THE PROBLEM

In the search for a new indicant of a disease D, doctors in a certain clinic selected 150 patients known to have the disease and 150 patients known not to have the disease. Here $\mathbf{D}$ is the event that a patient has the disease D, while $\mathbf{D}'$ is the event that a patient does not have the disease D. To each patient they apply a test obtaining a response E^+ for positive evidence or E^- for negative evidence. The results of the experiment are presented in Table 15.1.

A new patient comes to the clinic and is judged, by the doctors, to have the disease with probability 0.1. The doctors apply the same test to this patient

Influence Diagrams, Belief Nets and Decision Analysis
Edited by R. M. Oliver and J. Q. Smith

Table 15.1 Reslts of the clinical experiment

Patient's state	Patient's response		Sample size
	E_+	E_-	
D	60	90	150
D′	9	141	150

and obtain response E^+. How does this evidence change their probability that the patient has the disease? What would be this change if the response is E^-?

15.2 DIAGNOSTIC MODEL

To present a solution for this problem we define the following quantities which we think are the elements of the model:

1. The sensitivity of the test is $\pi = \Pr\{E^+ | \mathbf{D}\}$ and the specificity of the test is $\theta = \Pr\{E^- | \mathbf{D}'\}$.
2. The sampling quantities are $x|\pi \sim \text{bi}(150, \pi)$ and $y|\theta \sim \text{bi}(150, \theta)$. That is, x and y are binomial random quantities with parameters $(150, \pi)$ and $(150, \theta)$, respectively. Here x is the number of positive responses among the 150 patients having D and y is number of negative responses among the 150 patients not having D. We have observed $x = 60$ and $y = 141$.
3. The state of the new patient is

$$\delta = \begin{cases} 1 & \text{if the patient has the disease, } D \\ 0 & \text{otherwise.} \end{cases}$$

   The prior diagnostic probability is $\Pr\{\delta = 1\} = 0.1$. $\delta \sim \text{ber}(0.1)$ indicates that δ is a Bernoulli random quantity.
4. The result of the test for the new patient is

$$t = \begin{cases} 1 & \text{if a positive response obtains, i.e. } E^+ \\ 0 & \text{if a negative response obtains, i.e. } E^-. \end{cases}$$

   Note that $\Pr\{t = 1 | \delta = 1\} = \pi$ or that $\Pr\{t = 0 | \delta = 0\} = \theta$ if we judge, respectively, the new patient as we have judged the sample patients having D or the sample patients not having D. See Lindley and Novick (1981) for a complete discussion on exchangeability.
5. The posterior diagnostic probabilities, the object of the analysis, are $\Pr\{\delta = 0 | t = 0, x = 60, y = 141\}$ and $\Pr\{\delta = 1 | t = 1, x = 60, y = 141\}$. Note that the quantities on the right of the bar are observable and the ones on the left of the bar are the quantities of interest which at this stage are not

observable. The other quantities, π and θ, that are neither observable nor of interest, are eliminated during the analysis.

To construct the probabilistic influence diagram relating the nodes representing the above quantities we need to state the conditional independence relationships that we have judged to be relevant. The first and most important is $(x, \pi) \perp\!\!\!\perp (y, \theta)$; i.e. (x, π) and (y, θ) are independent. This is because (x, π) and (y, θ) are quantities related to two distinct and independent populations, **D** and **D**′. (We could think of two different urns having balls of two colors.) Since our interest is directed to a new patient (a different individual), his/her state δ is independent of the other patients in the sample. However, the response to the clinical test, t, given to a new patient depends on the value of π or θ and his/her state δ. With these restrictions in mind, Figure 15.1 presents our probabilistic influence diagram for the problem. To stress the fact that the clinical test is being given for the first time, we judge π and θ to have independent uniform distributions in the unit interval. Recall that the uniform density in the interval $(0, 1)$ is the beta density with parameters $a = b = 1$. The beta distribution with parameters a and b is denoted by $\text{Be}(a, b)$.

The diagram of Figure 15.1 (in our particular case $m = n = 150$) has four distinguished nodes. Three represent random quantities that have been observed and one represents the unknown quantity of interest, δ. The remaining nodes, π and θ, are the modeling parameters (i.e. are neither observable nor of interest) and must be eliminated. The directions of the arcs are also determined by the problem. The sample results, x and y, depend, respectively, on the chances of positive and negative responses, namely π and θ, in their respective populations **D** and **D**′. Analogously, the response of the new patient, t, depends on δ, the state of the patient, and on the accuracy of the test measured by π and θ.

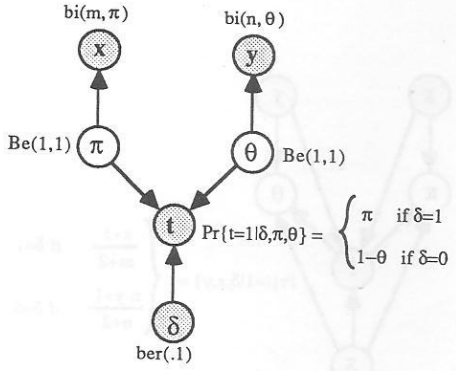


Figure 15.1 Probabilistic influence diagram for the diagnosis example.

15.3 THE 'IDEAL' SOLUTION

Figure 15.2 shows the probabilistic influence diagram after reversing arcs $[\pi, x]$ and $[\theta, y]$. After reversing these arcs, we use Bb$(m, 1, 1)$ and Bb$(n, 1, 1)$ to indicate that x and y are distributed as beta-binomial random quantities with parameters $(m, 1, 1)$ and $(n, 1, 1)$, respectively. (See Basu and Pereira, 1981, 1982 for a complete discussion on these distributions.) Arc reversal and node elimination, the diagram operations used here, are discussed by Barlow and Pereira (1987). After reversing arcs $[\pi, t]$ and $[\theta, t]$ we obtain the diagram of Figure 15.3. For simplicity, we give only the probability function of t since the distributions of x, y, and δ are given in Figure 15.2. Clearly the distributions of π and θ changed. Since π and θ are going to be eliminated (they are barren nodes in Figure 15.3), their probability functions do not appear in Figure 15.3. In fact we obtain:

1. $\pi|(\delta, t, x, y) \sim \pi|(\delta, t, x) \sim \text{Be}(1 + x + \delta t, 1 + m + \delta - x - \delta t)$; and

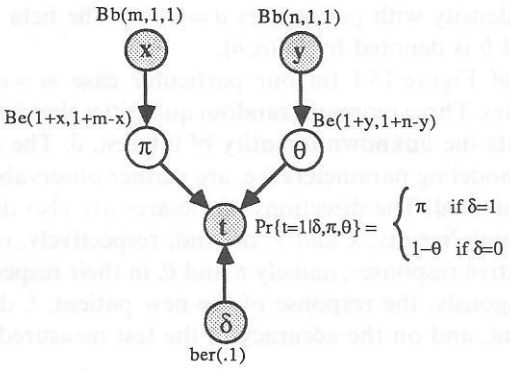


Figure 15.2 Probabilistic influence diagram after reversing arcs $[\pi, x]$ and $[\theta, y]$.

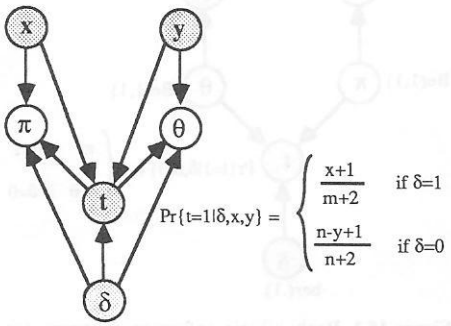


Figure 15.3 Probabilistic influence diagram.

2. $\theta|(\delta, t, x, y) \sim \theta|(\delta, t, y) \sim \text{Be}(1 + y + (1 - \delta)(1 - t),$
$1 + n + (1 - \delta) - y - (1 - \delta)(1 - t))$.

Althoughh these expressions look complicated they only reflect the fact that the new patient has to be added to the sample of $\mathbf{D}(\mathbf{D}')$ if $\delta = 1$ ($\delta = 0$) and, in this new sample, either $x(y)$ increases to $x + 1$ ($y + 1$) if his/her response was positive (negative) or $m - x$ increases to $m - x + 1$ in the case of a negative (positive) response. In addition, since the first (second) expression does not involve $y(x)$, we do not have to consider either arc $[x, \theta]$ or arc $[y, \pi]$ or arc $[\theta, \pi]$.

Figure 15.4 is our diagram after eliminating nodes π and θ. Since all the nodes are distinguished we did not shade them. The probabilistic influence diagram that permits us to evaluate the diagnostic probabilities for all possible values of the observable quantities, t, x, and y, is presented in Figure 15.5. The answer to our problem is given by the probability functions attached to node δ.

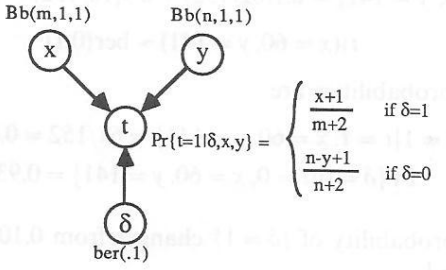


Figure 15.4 Probabilistic influence diagram after eliminating nodes π and θ.

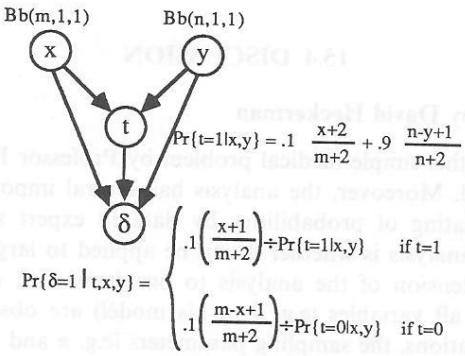


Figure 15.5 Probabilistic influence diagram after reversing node $[\delta, t]$.

Using now the experimental data displayed in Table 15.1 we obtain the following results:

1. The posterior distributions of π and θ are

$$\pi|(\delta, t, x = 60) \sim \begin{cases} \text{Be}(62, 91) & \text{if } \delta = 1, t = 1 \\ \text{Be}(61, 92) & \text{if } \delta = 1, t = 0 \quad \text{and} \\ \text{Be}(61, 91) & \text{if } \delta = 0. \end{cases}$$

$$\theta|(\delta, t, y = 141) \sim \begin{cases} \text{Be}(142, 10) & \text{if } \delta = 1 \\ \text{Be}(143, 10) & \text{if } \delta = 0, t = 0 \\ \text{Be}(142, 11) & \text{if } \delta = 0, t = 1. \end{cases}$$

2. The predictive distributions of x, y, and t are
   (a) $\Pr\{x = i\} = \Pr\{y = i\} = 1/151$, where $i = 0, 1, \ldots, 150$. That is, $x \sim y \sim \text{Bb}(150, 1, 1)$; and
   (b) $\Pr\{t = 1 | x = 60, y = 141\} = 0.1(62/152) + 0.9(10/152) = 15.2/152 = 0.1$, i.e.

$$t|(x = 60, y = 141) \sim \text{ber}(0.1).$$

3. The diagnostic probabilities are

$$\Pr\{\delta = 1 | t = 1, x = 60, y = 141\} = 61/152 = 0.40 \quad \text{and}$$
$$\Pr\{\delta = 0 | t = 0, x = 60, y = 141\} = 0.93.$$

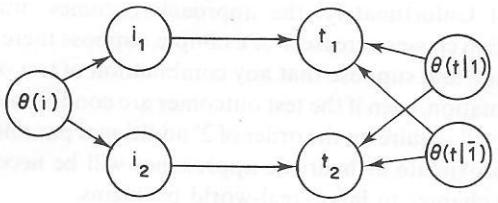
   Hence,
   (a) if $t = 1$ the probability of $\{\delta = 1\}$ changes from 0.10, *a priori*, to 0.40, *a posteriori*, and
   (b) if $t = 0$ the probability of $\{\delta = 0\}$ changes from 0.90, *a priori*, to 0.93, *a posteriori*.

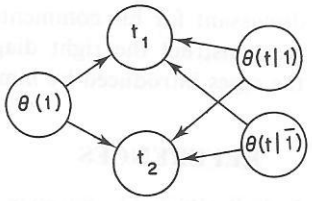
The fact that the change observed in (a) is bigger than that observed in (b) suggests that the diagnostic test in the study is more sensitive than it is specific.

15.4 DISCUSSION

15.4.1 Discussion by David Heckerman

The analysis of this simple medical problem by Professor Pereira is accurate and well presented. Moreover, the analysis has several important applications including the updating of probabilities by data in expert systems. My only concern with the analysis is whether it can be applied to larger, more realistic, problems. The extension of the analysis to problems with many variables is straightforward if all variables (e.g. δ in his model) are observed in repeated trials. In such situations, the sampling parameters (e.g. π and θ) can be updated independently. However, if one or more variables remain unobserved in certain trials, the parameters become dependent and updating becomes difficult.





To understand this point, consider a situation that is slightly more complicated than the one presented in the chapter. Suppose the outcome of the test depends on some intermediate variable, called i, where i can be absent ($i = 0$) or present ($i = 1$). For simplicity, consider only those patients with the disease D. An influence diagram for two patients is shown in Figure 15.6. The nodes subscripted with 1 and 2 represent the observable events for the first and second patient respectively. The nodes labeled $\theta(i)$, $\theta(t|i)$, and $\theta(t|\bar{i})$ represent the sampling parameters. If the intermediate variable for each of the two patients is observed, the sampling parameters can be updated independently. However, if the intermediate states remain unobserved, then the influence diagram reduces to the one shown in Figure 15.7. The parametes are no longer independent and updating becomes difficult when many patients are considered.

An approach to circumvent this difficulty is suggested by Ross Shachter in Chapter 14 of the volume. Suppose an additional parameter $\theta(t)$ is introduced, where θ is a deterministic function of the original three parameters:

$$\theta(t) = \theta(t|i)\theta(i) + \theta(t|\bar{i})(1 - \theta(i)).$$

Conditioned on this new parameter, observations for t, where i remainsunobserved, are independent. With the introduction of $\theta(t)$, inference becomes straightforward. When i is observed, either $\theta(t|i)$ or $\theta(t|\bar{i})$ is updated. When i is unobserved, $\theta(t)$ is updated. The probabilities of interest for a new patient can be computed from the parameters using a Monte Carlo approach, discretization, or other approximate approaches.

This method works well when there are a small number of observationclasses. (In the example above, there are three classes corresponding to i absent, i present,

and i unobserved.) Unfortunately, the approach becomes intractable as the number of observation classes ncrease. For example, suppose there are tests $t_1 \cdots t_n$, for which i is relevant, and suppose that any combination of test outcomes can be observed. In this situation, even if the test outcomes are conditionally independent given i, the method will require on the order of 2^n additional parameters. Therefore, it appears that approximate or heuristic approaches will be necessary to extend the analysis in this chapter to large, real-world problems.

15.4.2 Reply

I would like to thank the discussant for his comments. Also, I would like to say in a reply that, when we construct the right diagram for censored data problems, the solutions for the cases introduced by him could be obtained.

REFERENCES

Barlow, R. E. and Pereira, C. A. de B. (1987) *The Bayesian Operation and Probabilistic Influence Diagrams*. Berkeley, Department of Industrial Engineering and Operations Research, University of California, 50pp. (TR-ESRC 87-7)

Basu, D. and Pereira, C. A. de B. (1981) On Bayesian analysis of categorical survey data, *Bulletin of the International Statistical Institute*, Contributed Papers, **49**(2), 187–90.

Basu, D. and Pereira, C. A. de B. (1982) On the Bayesian analysis of categorical data: the problem of nonresponse, *Journal of Statistical Planning and Inference*, **6**(4), 345–62.

Lindley, D. V. and Novick, M. R. (1981) The role of exchangeability in inference, *Ann. Statist.*, **9**, 45–58.

Pereira, C. A. de B. and Pericchi, L. R. (1990) Analysis of diagnostability, *Applied Statistics*, **39** (1) (to appear).

Shachter, R. and Heckerman, D. (1987) Thinking backwards for knowledge acquisition, *AI Magazine*, **8** (Fall), 55–61.