CAUSALITY AND DECISION MAKING IN MEDICINE

Mary Randolph-Gips, Ph.D.
Assistant Professor, Computer Engineering and Systems Engineering
University of Houston - Clear Lake
e-mail: Randolph-gips@uhcl.edu

Ipek Bozkurt, Ph.D.
Assistant Professor, Engineering Management and Systems Engineering
University of Houston - Clear Lake
e-mail: Bozkurt@uhcl.edu

ABSTRACT

Statistics and causality play a major role in medical care today. This paper examines the fundamentals of Bayesian probability, Bayesian networks and causal networks. In particular, the effects of actions on causal networks are examined. Examples are given of applications of these techniques in the medical field.

1. INTRODUCTION

The medical field abounds with both trivial and life or death decisions. Although statistics are used in an attempt to optimize treatments, statistics reveal only passive, observational correlations. The model that most physicians use within their minds is causal. Causality is particularly important in the field of epidemiology (Rothman and Greenland 2005; De Vreese 2009). Causality has been used to examine such complex diseases as metabolic syndrome (Yudkin 2007), cancer (Greaves 2002), obesity (Oken 2009), and many of the other complex, multi-factorial illnesses that plague the modern world. Formal methods of causality can be used to refine experiment design and to allow proper adjustment for latent variables (Schisterman, Cole et al. 2009). Causality in the medical field is also quite important to the field of law (Sinclair 2010). To understand the issues involved with causality requires a background in Bayesian statistics and the causal framework developed by Judea Pearl (Pearl 2000). This paper will examine traditional Bayesian statistics for inference, and will expand into the area of causal modeling.

2. BAYES THEOREM

Bayes theorem gives the conditional probability of an event \(P(A|B)\) (the posterior probability) in terms of the prior probabilities of both A and B and the conditional probability of B given A, \(P(B|A)\). This is a very important concept, as it allows one to infer a conditional probability, \(P(A|B)\) that was not necessarily measured.
Conditional probability can also be defined in terms of the equation

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

This can be observed from the Venn diagram. The probability that A will be true is the fraction of B that is also A.

![Figure 1](image)

where \( P(AB) \) is the joint probability. If variables are independent, the joint probability is \( P(AB)=P(A)P(B) \).

This also implies that \( P(A|B)=P(A) \).

From the joint probability, the chain rule can be derived. In two variables it is

\[ P(AB) = P(A|B)P(B) \]

In three variables it is

\[ P(ABC) = P(A|BC)P(B|C)P(C) \]

or generalized to \( n \) variables it is

\[ P(X_1 \ldots X_n) = \prod_{i=1}^{n} P(X_i|X_1 \ldots X_{i-1}) \]

The order of the variables in the conditional probability terms is unimportant, as long as the conditional probabilities are available. Often, only conditional probabilities are available, so the chain rule is important as many statistics can be found from the joint probability. In particular, conditional probabilities which weren’t directly measured can be inferred using the joint and Bayes rule. For example, for three variables, the \( P(AB|C) \) can be inferred from the measured values \( P(A|BC), P(B|C) \) and \( P(C) \).
We can also eliminate variables in the probability by summing over the unwanted variables. We can calculate \( P(C) \) by summing \( P(ABC) \) over \( A \) and \( B \). This is called marginalizing.

\[
P(AB|C) = \frac{P(ABC)}{P(C)} = P(A|BC)P(B|C)
\]

These are important results, and when combined with the independence of model variables, are very powerful.

### 3. BAYES NETWORKS

A Bayes net is a graphical representation of the conditional probabilities. Every variable has a node to denote its probability. Measured conditional probabilities are shown by links or arcs between the nodes. The above equation \( P(ABC) = P(A|BC)P(B|C)P(C) \) can be shown as figure 1.

![Figure 1](image1.png)

A joint distribution can quickly grow to a large number of terms as the dimensionality of the probability space grows. For \( n \) binary valued probabilities, this would give \( 2^n \) terms. This is reduced greatly when terms are independent. Independence can be determined from the data, or can be inferred from domain knowledge. Bayes nets can be used to show that domain knowledge graphically. I will illustrate this with an example from my own past.

I had a physics professor who always liked to lecture late. Since the class period was from 4:30 - 5:20 pm, this often caused me to miss dinner at my scholarship hall. Sometimes there would be food left over, sometimes not. This was not just dependent on my time of arrival, but also upon the desirableness of the food that day. Also dependent on the quality of food served, was whether my roommates boyfriend would be visiting. This is shown in the diagram below. Domain knowledge is implicit in the independence of the professor, \( P \) and food served, \( F \). These variables are shown with no incoming links, and their probabilities are denoted by simply \( P(P) \) and \( P(F) \).
From this diagram we can calculate all sorts of interesting information. This can be found by first finding the joint using the chain rule.

\[ P(PLDFB) = P(D|PLFB)P(B|PLF)P(L|PF)P(F|P)P(P) \]

Using the independence relations encoded in the graph,

\[ P(D|PLFB) = P(D|LF) \]
\[ P(B|PLF) = P(B|F) \]
\[ P(L|PF) = P(L|P) \]
\[ P(F|P) = P(F) \]

Resulting in

\[ P(PLDFB) = P(D|LF)P(B|F)P(L|P)P(F)P(P) \]

From this the probability of all sorts of odd things can be calculated, such as the probability of the boyfriend given the professor, or even the professor given the boyfriend using Bayes rule and summing over the unwanted variables.

### 4. CAUSALITY NETWORKS

Thus far, Bayes nets have been discussed in terms of representing probability functions. They can also be thought of as representing causal information. The professor lecturing late caused me to miss dinner. This is shown by the directed arrows between the nodes. These networks
are called Directed Acyclic Graphs (DAG)s. In order to represent the temporal component of causality, no loops are allowed.

An implication of the directed nature of the causal diagrams is the conditional independence of the nodes. There are three types of connecting nodes.

The first is a chain node, $A \rightarrow B \rightarrow C$. If we know the value of $B$, or condition on $B$, then it doesn’t matter what $A$ is as far as $C$ is concerned. The dependence of $B$ upon $A$ has been removed from consideration. Thus, $A$ and $C$ are independent, conditioned on $B$. $A$ and $C$ are not independent if we don’t know the value of $B$. An example of a chain node is the sequence, PLD in the late dinner example. If we know that the class ran late, it doesn’t matter who the professor was, there is still a greater chance that no dinner was left.

The second type of node is the diverging node, $B \leftarrow A \rightarrow C$. If we condition on $A$, then $B$ and $C$ are conditionally independent. However, if we do not condition on $A$, then $B$ and $C$ are not independent. Look at the nodes D,F and B in the dinner diagram. If we know that the boyfriend is present, then the probability of yummy food is increased, which then increases the probability of no dinner being left. So these values are dependent if we don’t know the value of food. If we do know the value of food, then D and B are independent.

The third type of node is the trickiest. A converging node, $A \rightarrow B \leftarrow C$ will have the opposite behavior of the other two node types upon conditioning. $A$ and $C$ are independent when we don’t know $B$. If we condition on $B$, then $A$ and $C$ are rendered dependent. Consider nodes L,D and F from the dinner example. If we know that there is dinner left, the fact that the food was good increases the chance that the class did not run late.

4.1 Actions in Causality Networks

Causal diagrams also allow the possibility of actions upon the graph. Tired of always missing dinner, my classmates and I decided to take action. We turned the clock ten minutes ahead. With this action, we tried to force the class to end on time, severing the link between Professor and Late in the causal diagram. Pearl (Pearl 2010) uses the do(.) operator to denote actions taken upon a variable. Using the do notation, we do(L)=0. This action is fundamentally different from simply observing the value of a parameter. The graph is now fundamentally changed, although we can still use the conditional probabilities calculated with the intact graph. Let’s add some nodes the graph to make this more clear. There is a certain probability that I would go out for pizza, given whether there was any food left, and the quality of food. If I go out for pizza, then the chances of finishing my homework decrease. Thus, whether or not I complete my homework will depend on both my professor that day and on the quality of food. If I merely
observed that there was some dinner left to eat, as in then the probability of homework completion would still be related to the food through two paths as in figure 3a.

Now back to the original example. All was going well after we set the clock ahead. Dr. Culvahouse showed up at 4:30, got a little frazzled at thinking he had shown up late, and proceeded to lecture. At what he thought was 5:30, he started wrapping up the lecture. Just then the campus wide end of class whistle blew. Dr. Culvahouse realized that the clock was wrong and lectured for ten more minutes. Our plan was foiled by an unconsidered, unmeasured, latent variable. Bayes and causal networks are well set up to deal with latent variables. Latent variables can be explicitly included in the diagram, with unmeasured probabilities. This is particularly important in the medical field, where the complexity of the human body means that latent variables are endemic.

5. APPLICATIONS IN MEDICAL FIELDS

Bayesian networks have a long history in the medical field. An early application was the Pathfinder system to diagnose lymph node disease (Heckerman, Horvitz et al. 1992). Gene
regulation has in particular been a ripe field. Finding splice sites is the focus of many articles (Cai, Delcher et al. 2000; Degroeve, Saeys et al. 2005). Cellular networks have also been modeled using Bayes techniques. (Friedman, Linial et al. 2000; Friedman 2004) Causal graphs have been used to reduce bias in medical research. (Shrier and Platt 2008)

Causal techniques can be used to develop new techniques in data mining. In particular, techniques can be developed to utilize both quantitative and qualitative information using causal graphs.

Good references on Causal and Bayesian networks can be found in (Charniak 1991; Pearl 2000; Pearl 2003).

6. CONCLUSIONS

This paper reviewed some of the concepts of probability and causal theory. Conditional independence and effects of actions on diagrams were illustrated. These concepts form the basis of the applications to medical causality.

REFERENCES