
The Noisy-Logical Distribution and its Application to Causal Inference

Alan Yuille

Department of Statistics
University of California at Los Angeles
Los Angeles, CA 90095
yuille@stat.ucla.edu

HongJing Lu

Department of Psychology
University of California at Los Angeles
Los Angeles, CA 90095

To appear in Advances in Neural Information Processing Systems 20. Cambridge, MA: MIT Press. A Bradford Book. To appear 2008.

The Noisy-Logical Distribution and its Application to Causal Inference

Abstract

We describe a novel noisy-logical distribution for representing the distribution of a binary output variable conditioned on multiple binary input variables. The distribution is represented in terms of noisy-or's and noisy-and-not's of *causal features* which are conjunctions of the binary inputs. The standard noisy-or and noisy-and-not models, used in causal reasoning and artificial intelligence, are special cases of the noisy-logical distribution. We prove that the noisy-logical distribution is complete in the sense that it can represent all conditional distributions provided a sufficient number of causal factors are used. We illustrate the noisy-logical distribution by showing that it can account for new experimental findings on how humans perform causal reasoning in more complex contexts. Finally, we speculate on the use of the noisy-logical distribution for causal reasoning and artificial intelligence.

1 Introduction

The noisy-or and noisy-and-not conditional probability distributions are frequently studied in cognitive science for modeling causal reasoning [1], [2],[3] and are also used as probabilistic models for artificial intelligence [4]. It has been shown, for example, that human judgments of the power of causal cues in certain experiments [1] can be interpreted in terms of maximum likelihood estimation and model selection using these types of models [3].

But the noisy-or and noisy-and-not distributions are limited in the sense that they can only represent a restricted set of all possible conditional distributions. This restriction is sometimes an advantage because there may not be sufficient data to determine the full conditional distribution. Nevertheless it would be better to have a representation that can expand to represent the full conditional distribution, if sufficient data is available, but can be reduced to simpler forms (e.g. standard noisy-or) if only limited data is available.

This motivates us to define the noisy-logical distribution. This is defined in terms of noisy-or's and noisy-and-not's of *causal features* which are conjunctions of the basic input variables. (The use of conjunctive features to extend the noisy-or and noisy-and-not models was proposed in [2]). By restricting the choice of causal features we can obtain the standard noisy-or and noisy-and-not models. We prove that the noisy-logical distribution is complete in the sense that it can represent any conditional distribution provided we use all the causal features. Overall, it gives a distribution whose complexity can be adjusted by restricting the set of causal features.

To illustrate the noisy-logical distribution we apply it to modeling some recent human experiments on causal reasoning in complex environments [5]. We show that noisy-logical distributions involving causal factors are able to account for human performance. By contrast, an alternative linear model gives predictions which are the opposite of the observed trends in human causal judgments. Section (2) presents the noisy-logical distribution for the case with two input causes (the case commonly studied in causal reasoning). In section (3) we specify the full noisy-logical distribution and we prove its completeness in section (4). Section (5) illustrates the noisy-logical distribution by showing that it accounts for recent experimental findings in causal reasoning.

2 The Case with $N = 2$ causes

In this section we study the simple case when the binary output effect E depends only on two binary-valued causes C_1, C_2 . This covers most of the work reported in the cognitive science literature [1],[3]. In this case, the probability distribution is specified by the four numbers $P(E = 1|C_1, C_2)$, for $C_1 \in \{0, 1\}$, $C_2 \in \{0, 1\}$.

To define the noisy-logical distribution over two variables $P(E = 1|C_1, C_2)$, we introduce three concepts. Firstly, we define four binary-valued causal features $\Psi_0(\cdot), \Psi_1(\cdot), \Psi_2(\cdot), \Psi_3(\cdot)$ which are functions of the input state $\vec{C} = (C_1, C_2)$. They are defined by $\Psi_0(\vec{C}) = 1, \Psi_1(\vec{C}) = C_1, \Psi_2(\vec{C}) = C_2, \Psi_3(\vec{C}) = C_1 \wedge C_2$, where \wedge denotes logical-and operation (i.e. $C_1 \wedge C_2 = 1$ if $C_1 = C_2 = 1$ and $C_1 \wedge C_2 = 0$ otherwise). $\Psi_3(\vec{C})$ is the conjunction of C_1 and C_2 . Secondly, we introduce binary-valued hidden states E_0, E_1, E_2, E_3 which are caused by the corresponding features $\Psi_0, \Psi_1, \Psi_2, \Psi_3$. We define $P(E_i = 1|\Psi_i; \omega_i) = \omega_i \Psi_i$ with $\omega_i \in [0, 1]$, for $i = 1, \dots, 4$ with $\vec{\omega} = (\omega_1, \omega_2, \omega_3, \omega_4)$. Thirdly, we define the output effect E to be a logical combination of the states E_0, E_1, E_2, E_3 which we write in form $\delta_{E, f(E_0, E_1, E_2, E_3)}$, where $f(\cdot, \cdot, \cdot, \cdot)$ is a logic function which is formed by a combination of three logic operations *AND*, *OR*, *NOT*. This induces the *noisy-logical distribution* $P_{nl}(E|\vec{C}; \vec{\omega}) = \sum_{E_0, \dots, E_3} \delta_{E, f(E_0, E_1, E_2, E_3)} \prod_{i=0}^3 P(E_i|\Psi_i(\vec{C}); \omega_i)$.

The noisy-logical distribution is characterized by the parameters $\omega_0, \dots, \omega_3$ and the choice of the logic function $f(\cdot, \cdot, \cdot, \cdot)$. We can represent the distribution by a circuit diagram where the output E is a logical function of the hidden states E_0, \dots, E_3 and each state is caused probabilistically by the corresponding causal features Ψ_0, \dots, Ψ_3 , as shown in Figure (1).

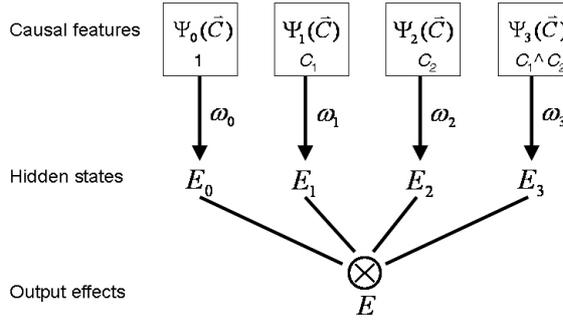


Figure 1: Circuit diagram in the case with $N = 2$ causes.

The noisy-logical distribution includes the commonly known distributions, noisy-or and noisy-and-not, as special cases. To obtain the noisy-or, we set $E = E_1 \vee E_2$ (i.e. $E_1 \vee E_2 = 0$ if $E_1 = E_2 = 0$ and $E_1 \wedge E_2 = 1$ otherwise). A simple calculation shows that the noisy-logical distribution reduces to the noisy-or $P_{nor}(E|C_1, C_2; \omega_1, \omega_2)$ [4], [7]:

$$\begin{aligned} P_{nl}(E = 1|C_1, C_2; \omega_1, \omega_2) &= \sum_{E_1, E_2} \delta_{1, E_1 \vee E_2} P(E_1|\Psi_1(\vec{C}); \omega_1) P(E_2|\Psi_2(\vec{C}); \omega_2) \\ &= \omega_1 C_1 (1 - \omega_2 C_2) + (1 - \omega_1 C_1) \omega_2 C_2 + \omega_1 \omega_2 C_1 C_2 \\ &= \omega_1 C_1 + \omega_2 C_2 - \omega_1 \omega_2 C_1 C_2 = P_{nor}(E = 1|C_1, C_2; \omega_1, \omega_2) \quad (1) \end{aligned}$$

To obtain the noisy-and-not, we set $E = E_1 \wedge \neg E_2$ (i.e. $E_1 \wedge \neg E_2 = 1$ if $E_1 = 1, E_2 = 0$ and $E_1 \wedge \neg E_2 = 0$ otherwise). The noisy-logical distribution reduces to the noisy-and-not $P_{n-and-not}(E|C_1, C_2; \omega_1, \omega_2)$ [4],[7]:

$$\begin{aligned} P_{nl}(E = 1|C_1, C_2; \omega_1, \omega_2) &= \sum_{E_1, E_2} \delta_{1, E_1 \wedge \neg E_2} P(E_1|\Psi_1(\vec{C}); \omega_1) P(E_2|\Psi_2(\vec{C}); \omega_2) \\ &= \omega_1 C_1 \{1 - \omega_2 C_2\} = P_{n-and-not}(E = 1|C_1, C_2; \omega_1, \omega_2). \quad (2) \end{aligned}$$

We claim that noisy-logical distributions of this form can represent any conditional distribution $P(E|\vec{C})$. The logical function $f(E_0, E_1, E_2, E_3)$ will be expressed as a combination of logic operations *AND*-*NOT*, *OR*. The parameters of the distribution are given by $\omega_0, \omega_1, \omega_2, \omega_3$.

The proof of this claim will be given for the general case in the next section. To get some insight, we consider the special case where we only know the values $P(E|C_1 = 1, C_2 = 0)$ and $P(E|C_1 = 1, C_2 = 1)$. This situation is studied in cognitive science where C_1 is considered to be a background cause which always takes value 1, see [7] [6]. In this case, the only causal features are considered, $\Psi_1(\vec{C}) = C_1$ and $\Psi_2(\vec{C}) = C_2$.

Result. *The noisy-or and the noisy-and-not models, given by equations (1,2) are sufficient to fit any values of $P(E = 1|1, 0)$ and $P(E = 1|1, 1)$. (In this section we use $P(E = 1|1, 0)$ to denote $P(E = 1|C_1 = 1, C_2 = 0)$ and use $P(E = 1|1, 1)$ to denote $P(E = 1|C_1 = 1, C_2 = 1)$.) The noisy-or and noisy-and-not fit the cases when $P(E = 1|1, 1) \geq P(E = 1|1, 0)$ and $P(E = 1|1, 1) \leq P(E = 1|1, 0)$ respectively. (Equivalently, C_2 is respectively a generative or preventative cause).*

Proof. We can fit both the noisy-or and noisy-and-not models to $P(E|1, 0)$ by setting $\omega_1 = P(E = 1|1, 0)$, so it remains to fit the models to $P(E|1, 1)$. There are three cases to consider: (i) $P(E = 1|1, 1) > P(E = 1|1, 0)$, (ii) $P(E = 1|1, 1) < P(E = 1|1, 0)$, and (iii) $P(E = 1|1, 1) = P(E = 1|1, 0)$. It follows directly from equations (1,2) that $P_{nor}(E = 1|1, 1) \geq P_{nor}(E = 1|1, 0)$ and $P_{n-and-not}(E = 1|1, 1) \leq P_{n-and-not}(E = 1|1, 0)$ with equality only if $P(E = 1|1, 1) = P(E = 1|1, 0)$. Hence we must fit a noisy-or and a noisy-and-not model to cases (i) and (ii) respectively. For case (i), this requires solving $P(E = 1|1, 1) = \omega_1 + \omega_2 - \omega_1\omega_2$ to obtain $\omega_2 = \{P(E = 1|1, 1) - P(E = 1|1, 0)\}/\{1 - P(E = 1|1, 0)\}$ (note that the condition $P(E = 1|1, 1) > P(E = 1|1, 0)$ ensures that $\omega_2 \in [0, 1]$). For case (ii), we must solve $P(E = 1|1, 1) = \omega_1 - \omega_1\omega_2$ which gives $\omega_2 = \{P(E = 1|1, 0) - P(E = 1|1, 1)\}/P(E = 1|1, 0)$ (the condition $P(E = 1|1, 1) < P(E = 1|1, 0)$ ensures that $\omega_2 \in [0, 1]$). For case (iii), we can fit either model by setting $\omega_2 = 0$.

To anticipate terminology used in the general proof. We define $E^1 = E_1$, and fit the data $P(E|1, 0)$ by setting $E = E^1$ and $P(E_1|\Psi_1; \omega_1) = \omega_1\Psi_1$, with $\omega_1 = P(E = 1|1, 0)$. If $P(E = 1|1, 1) > P(E = 1|1, 0)$, we define $E = E^1 \vee E_2$ and $P(E_2 = 1|\Psi_2; \omega_2) = \omega_2\Psi_2$, with $\omega_2 = \{P(E = 1|1, 1) - P(E = 1|1, 0)\}/\{1 - P(E = 1|1, 0)\}$. If $P(E = 1|1, 1) < P(E = 1|1, 0)$, then we set $E = E^1 \wedge \neg E_2$ and $P(E_2 = 1|\Psi_2; \omega_2) = \omega_2\Psi_2$, with $\omega_2 = \{P(E = 1|1, 0) - P(E = 1|1, 1)\}/P(E = 1|1, 0)$. If $P(E = 1|1, 1) = P(E = 1|1, 0)$, then we only define $E = E^1$.

3 The Noisy-Logical Distribution for N causes

We next consider representing probability distributions of form $P(E|\vec{C})$, where $E \in \{0, 1\}$ and $\vec{C} = (C_1, \dots, C_N)$ where $C_i \in \{0, 1\}$, $\forall i = 1, \dots, N$. These distributions can be characterized by the values of $P(E = 1|\vec{C})$ for all possible 2^N values of \vec{C} .

We define the set of 2^N binary-valued causal features $\{\Psi_i(\vec{C}) : i = 0, \dots, 2^N - 1\}$. These features are ordered so that $\Psi_0(\vec{C}) = 1$, $\Psi_i(\vec{C}) = C_i : i = 1, \dots, N$, $\Psi_{N+1}(\vec{C}) = C_1 \wedge C_2$ is the conjunction of C_1 and C_2 , and so on. The feature $\Psi(\vec{C}) = C_a \wedge C_b \wedge \dots \wedge C_g$ will take value 1 if $C_a = C_b = \dots = C_g = 1$ and value 0 otherwise.

We define binary variables $\{E_i : i = 0, \dots, 2^N - 1\}$ which are related to the causal features $\{\Psi_i : i = 0, \dots, 2^N - 1\}$ by distributions $P(E_i = 1|\Psi_i; \omega_i) = \omega_i\Psi_i$, specified by parameters $\{\omega_i : i = 0, \dots, 2^N - 1\}$. Then we define the output variable E to be a logical (i.e. deterministic) function of the $\{E_i : i = 0, \dots, 2^N - 1\}$. This can be thought of as a circuit diagram. In particular, we define $E = f(E_0, \dots, E_{2^N-1}) = (((((E_1 \otimes E_2) \otimes E_3) \otimes E_4 \dots)$ where $E_1 \otimes E_2$ can be $E_1 \vee E_2$ or $E_1 \wedge \neg E_2$ (where $\neg E$ means logical negation). This means that we can define variables E^i recursively, by $E^{i+1} = E^i \otimes E_i$ with $E^1 = E_1$. This gives the general *noisy-logical distribution*, as shown in Figure (2).

$$P(E = 1|\vec{C}; \vec{\omega}) = \sum_{\vec{E}} \delta_{E, f(E_0, \dots, E_{2^N-1})} \prod_{i=0}^{2^N-1} P(E_i = 1|\Psi_i; \omega_i). \quad (3)$$

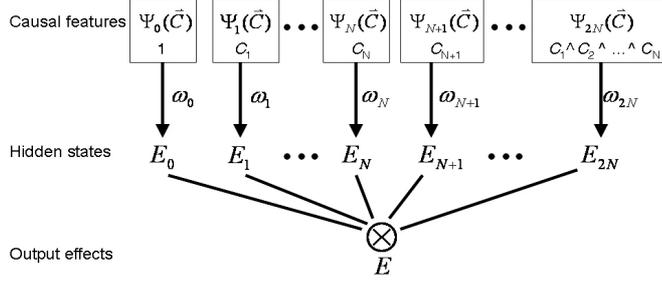


Figure 2: Circuit diagram in the case with N causes.

4 The Completeness Result

This section proves that the noisy-logical distribution is capable of representing any conditional distribution. This is the main theoretical result of this paper.

Result *We can represent any conditional distribution $P(E|\vec{C})$ defined on binary variables in terms of a noisy logical distribution given by equation (3).*

Proof. The proof is constructive. We show that any distribution $P(E|\vec{C})$ can be expressed as a noisy-logical distribution. The strategy is to use induction to build a noisy-logical distribution which agrees with $P(E|\vec{C})$ for all values of \vec{C} .

We order the states $\vec{C}_0, \dots, \vec{C}_{2^N-1}$. This ordering must obey $\Psi_i(\vec{C}_i) = 1$ and $\Psi_i(\vec{C}_j) = 0, \forall j < i$. This ordering can be obtained by setting $\vec{C}_0 = (0, \dots, 0)$, then selecting the terms with a single conjunction (i.e. only one C_i is non-zero), then those with two conjunctions (i.e. two C_i 's are non-zero), then with three conjunctions, and so on.

We start the induction using feature $\Psi_0(\vec{C}) = 1$. Set $E^0 = E_0$ and $\omega_0 = P(E|0, \dots, 0)$. Then $P(E^0|\vec{C}_0; \omega_0) = P(E|\vec{C}_0)$, so the noisy-logical distribution fits the data for input \vec{C}_0 .

Now proceed by induction. Assume we have E^M and $\omega_0, \dots, \omega_M$ such that $P(E^M = 1|\vec{C}_i; \omega_0, \dots, \omega_M) = P(E = 1|\vec{C}_i)$, for $i = 0, \dots, M$. (E^M is a logical function of E_0, \dots, E_M). There are three cases to consider which are analogous to the cases considered in the section with two causes.

Case 1. If $P(E = 1|\vec{C}_{M+1}) > P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M)$ we need $\Psi_{M+1}(\vec{C})$ to be a generative feature. Set $E^{M+1} = E^M \vee E_{M+1}$ with $P(E_{M+1} = 1|\Psi_{M+1}; \omega_{M+1}) = \omega_{M+1}\Psi_{M+1}$. Then we obtain:

$$\begin{aligned} P(E^{M+1} = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_{M+1}) &= P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M) + P(E_{M+1}|\Psi_{M+1}(\vec{C}); \omega_{M+1}) \\ &\quad - P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M)P(E_{M+1} = 1|\Psi_{M+1}(\vec{C}); \omega_{M+1}) \\ &= P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M) + \omega_{M+1}\Psi_{M+1}(\vec{C}) - P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M)\omega_{M+1}\Psi_{M+1}(\vec{C}). \end{aligned} \quad (4)$$

In particular, we see that $P(E^{M+1} = 1|\vec{C}_i; \omega_0, \dots, \omega_{M+1}) = P(E^M = 1|\vec{C}_i; \omega_0, \dots, \omega_M) = P(E = 1|\vec{C}_i)$ for $i < M + 1$ (using $\Psi_{M+1}(\vec{C}_i) = 0, \forall i < M + 1$). To determine the value of ω_{M+1} , we must solve $P(E = 1|\vec{C}_{M+1}) = P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M) + \omega_{M+1} - P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M)\omega_{M+1}$ (using $\Psi_{M+1}(\vec{C}_{M+1}) = 1$). This gives $\omega_{M+1} = \{P(E = 1|\vec{C}_{M+1}) - P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M)\} / \{1 - P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_{M+1})\}$ (the conditions ensure that $\omega_{M+1} \in [0, 1]$).

Case 2. If $P(E = 1|\vec{C}_{M+1}) < P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M)$ we need $\Psi_{M+1}(\vec{C})$ to be a preventative feature. Set $E^{M+1} = E^M \wedge \neg E_{M+1}$ with $P(E_{M+1} = 1|\Psi_{M+1}; \omega_{M+1}) = \omega_{M+1}\Psi_{M+1}$. Then we obtain:

$$P(E^{M+1} = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_{M+1}) = P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M)\{1 - \omega_{M+1}\Psi_{M+1}(\vec{C})\}. \quad (5)$$

As for the first case, $P(E^{M+1} = 1|\vec{C}_i; \omega_0, \dots, \omega_{M+1}) = P(E^M = 1|\vec{C}_i; \omega_0, \dots, \omega_M) = P(E = 1|\vec{C}_i)$ for $i < M + 1$ (because $\Psi_{M+1}(\vec{C}_i) = 0, \forall i < M + 1$). To determine the value of ω_{M+1} we must solve $P(E = 1|\vec{C}_{M+1}) = P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M)\{1 - \omega_{M+1}\}$ (using $\Psi_{M+1}(\vec{C}_{M+1}) = 1$). This gives $\omega_{M+1} = \{P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M) - P(E = 1|\vec{C}_{M+1})\}/P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M)$ (the conditions ensure that $\omega_{M+1} \in [0, 1]$).

Case 3. If $P(E = 1|\vec{C}_{M+1}) = P(E^M = 1|\vec{C}_{M+1}; \omega_0, \dots, \omega_M)$, then we don't need to do anything.

5 Cognitive Science Human Experiments

We illustrate the noisy-logical distribution by applying it to model two new experimental studies [5]. These studies involve several cues (i.e. more than the standard two causal cues used in typical studies [1], [2],[3]). In addition, they can also involve conjunctions of basic causal cues. We compare the predictions of the noisy-logical distribution to the experimental findings and, for comparison, to the predictions of the linear model (a natural extension of the linear model also studied in [1],[3])

5.1 Experiment I: Multiple Causes

We tested the causal representation proposed in this paper by using the model to simulate human causal judgments in complex situations involving multiple causes and causal interactions. Liljeholm and Cheng [5] conducted innovative experiments on human causal judgment in complex environments. In Experiment 1 of [5], the cover story involved a set of allergy patients who either did or did not have a headache (E), and either had or had not received allergy medicines A, B. Participants were informed that two independent studies were conducted in different labs using different patient groups. In the first study patients were administered medicine A, whereas in the second study patients were administered both medicines A and B. The query concerned whether medicine B caused headache as a side effect. A simultaneous presentation format [6] was used to display specific contingency conditions used in each of the studies described to the participants in Experiment 1.

In this experimental setup, the output effect (E) is the headache; B_1 and B_2 is defined as background cause in the first and the second study, respectively; C_1 and C_2 correspond to two candidate causes, Medicine A and B. We use $P(E = 1|B_1 = 1, C_1 = 0, C_2 = 0)$ to indicate number of patients with headache out of number who had not received any medicine in study 1; $P(E = 1|B_1 = 1, C_1 = 1, C_2 = 0)$ indicate number of patients with headache out of number who had only received medicine A in study 1; $P(E = 1|B_2 = 1, C_1 = 0, C_2 = 0)$ indicate number of patients with headache out of number who had not received any medicine in study 2; $P(E = 1|B_2 = 1, C_1 = 1, C_2 = 1)$ indicate number of patients with headache out of number who had received both medicine A and B in study 2. Two conditions were included in Experiment 1. (a)Power-constant condition: in study 1, $P(E = 1|B_1 = 1, C_1 = 0, C_2 = 0) = 16/24, P(E = 1|B_1 = 1, C_1 = 1, C_2 = 0) = 22/24$; in study 2, $P(E = 1|B_2 = 1, C_1 = 0, C_2 = 0) = 0/24, P(E = 1|B_2 = 1, C_1 = 1, C_2 = 1) = 18/24$. (b) ΔP -constant condition: in study 1, $P(E = 1|B_1 = 1, C_1 = 0, C_2 = 0) = 0/24, P(E = 1|B_1 = 1, C_1 = 1, C_2 = 0) = 6/24$; in study 2, $P(E = 1|B_2 = 1, C_1 = 0, C_2 = 0) = 16/24, P(E = 1|B_2 = 1, C_1 = 1, C_2 = 1) = 22/24$. The query was, "Based on the information from BOTH studies, what is your best bet on whether or not medicine B causes headache?" The dependent measure was the proportion of participants who judged that Medicine B caused headache. Data were based on responses from a total of 50 undergraduates (25 participants in each condition).

A Bayesian model can be formalized to infer causal structure, as shown in Figure (3), by assessing whether a causal link exists between medicine B (C_2) and headaches (E). The decision variable is obtained from the log likelihood ratio of Graphs 2 and 1, $\log \frac{P(D|Graph2)}{P(D|Graph1)}$, termed *causal support* [3]. The likelihoods on graphs are computed by integrating out the unknown causal strengths w_1, w_2, w_3 , and w_4 , which are parameters in the range 0,1 associated with background cause in the first study B_1 , medicine A C_1 , background cause in the second study B_2 and medicine B C_2 , respectively. Human participants were informed that the two candidate causes, medicine A and B, generates the side effect of headache in the experiment. Furthermore, there was no explicit statement indicating the existence of a possible causal interaction between medicines A and B. We assume a noisy-or

distribution in both causal graphs, $E = E_1 \vee E_2$ in graph 1 and $E = E_2 \vee E_3 \vee E_4$ in graph 2

$$P(D|Graph2) = \int_0^1 P(D|w_1, w_2, w_3, w_4, Graph2)P(w_1, w_2, w_3, w_4|Graph2)dw_1dw_2dw_3dw_4, \quad (6)$$

$$P(D|Graph1) = \int_0^1 P(D|w_1, w_2, w_3, Graph1)P(w_1, w_2, w_3|Graph1)dw_1dw_2dw_3, \quad (7)$$

where $P(w_1, w_2, w_3, w_4|Graph2)$ and $P(w_1, w_2, w_3|Graph1)$ are assumed to follow uniform distributions. $P(D|w_1, w_2, w_3, w_4, Graph2)$ and $P(D|w_1, w_2, w_3, Graph1)$ are the likelihood probabilities of the observed data given specified causal strengths ω s and structures as shown Figure (3). Likelihood distributions are modeled using the noisy-or representation,

$$P(D|w_1, w_2, w_3, w_4, Graph2) = \sum_{\bar{E}} \delta_{E, (E_1 \vee E_2, E_2 \vee E_3 \vee E_4)} \prod_{i=1}^4 P(E_i = 1 | \Psi_i; \omega_i), \quad (8)$$

$$P(D|w_1, w_2, w_3, Graph1) = \sum_{\bar{E}} \delta_{E, (E_1 \vee E_2, E_2 \vee E_3)} \prod_{i=1}^3 P(E_i = 1 | \Psi_i; \omega_i). \quad (9)$$

5.2 Results

The specific contingency conditions in each study are shown in the caption of Figure (3). The left plot in Figure (3) depicts the proportion of participants who responded " Yes " to the query whether medicine B causes headaches in the two conditions, Power-constant condition (black bars) and ΔP -constant condition (white bars). In the first condition, causal powers remained the same in the two studies (contexts) but ΔP varied. In the second condition, ΔP was constant in the two studies but causal powers varied. The human data clearly indicate that change across contexts is largely determined by variation in causal power rather than in ΔP .

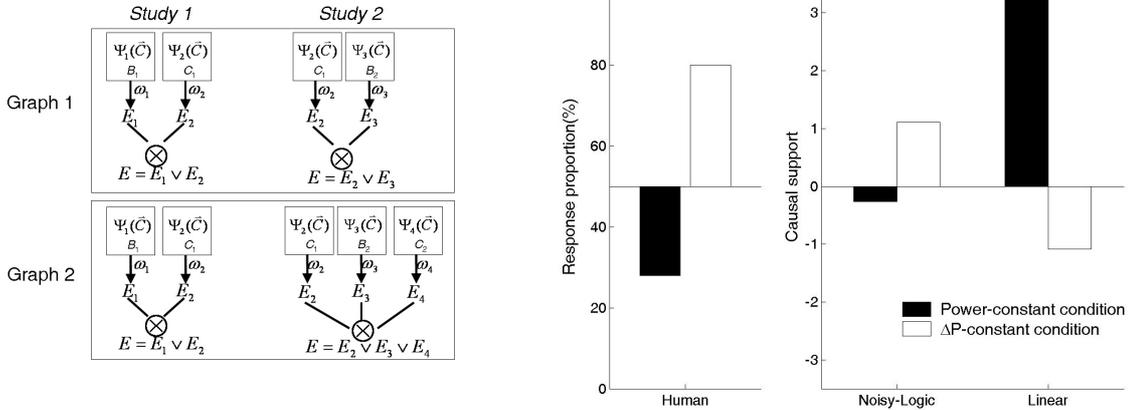


Figure 3: Model diagram and results in Experiment I. Left panel: Causal structure with multiple causes. Right panel: Results in a causal link judgment. Two experimental conditions, Power-constant condition and ΔP -constant condition. Left plot depicts human data (proportion of participants who responded Yes indicating medicine B caused headaches). Right plot depicts model predictions of causal support, including the noisy-logical model and the linear model.

The right plot in Figure (3) depicts model predictions with respect to causal support for medicine B (cause C_2 in Figure (3)) in the two experimental conditions. The noisy-logical model is based on the model proposed in the previous section. An alternative model is the linear model, which uses the same causal features as the noisy-logical model but rather than the logical function employs an additive function with the constraint that conditional probability lies within the range of 0 to 1

[8]. By comparing the two models, we find that human trend (higher rating for medicine B as a cause of headache in the ΔP -constant condition) is predicted by the Bayesian model with Noisy-or representation using *causal factors*, but not with the linear representation using the same causal factors.

5.3 Human Experiment II: Causal Interaction

Liljeholm and Cheng [5] developed another novel experimental paradigm to investigate causal interaction. The experimental design was identical to that used in Experiment 1, except that participants were presented with three studies in which only one medicine (A) was tested. Participants were asked to judge whether that medicine A interacts with background cause that might vary across the three studies. We define background cause as B_1, B_2, B_3 in the three studies, and candidate cause as C_1 for medicine A. Two conditions were used in this experiment, (a) Power-constant condition: in study 1, $P(E = 1|B_1 = 1, C_1 = 0) = 16/24$, $P(E = 1|B_1 = 1, C_1 = 1) = 22/24$; in study 2, $P(E = 1|B_2 = 1, C_1 = 0) = 8/24$, $P(E = 1|B_2 = 1, C_1 = 1) = 20/24$; in study 3, $P(E = 1|B_3 = 1, C_1 = 0) = 0/24$, $P(E = 1|B_3 = 1, C_1 = 1) = 18/24$; (b) Power-varying condition: in study 1, $P(E = 1|B_1 = 1, C_1 = 0) = 0/24$, $P(E = 1|B_1 = 1, C_1 = 1) = 6/24$; in study 2, $P(E = 1|B_2 = 1, C_1 = 0) = 0/24$, $P(E = 1|B_2 = 1, C_1 = 1) = 12/24$; in study 3, $P(E = 1|B_3 = 1, C_1 = 0) = 0/24$, $P(E = 1|B_3 = 1, C_1 = 1) = 18/24$.

A Bayesian model can be formalized to assess whether a causal link exists between the conjunctive cause $B \wedge C_1$ and E as shown in Figure (4). The decision variable in the model is causal support as defined in equations (6,7). However, the causal direction of conjunctive cause $B \wedge C_1$ is unknown, which could be generative (e.g. produce headaches) or preventive (e.g. prevent headaches). Accordingly, the noisy-logical distribution in this experiment includes two possible logical functions as shown in graph 2 of Figure (4).

5.4 Results

Figure (4) shows human and model performance in the two conditions, Power-constant condition and Power-varying condition, with respect to causal support for causal interaction between candidate cause and background cause ($B \wedge C_1$). The noisy-logical model correctly predicts that human trend, in which most participants responded No to the interaction query in the Power-constant condition, and most participants responded Yes in the Power-varying condition. In contrast, the linear model predicts that most participants responded Yes in both experimental conditions, which is opposite to human trend.

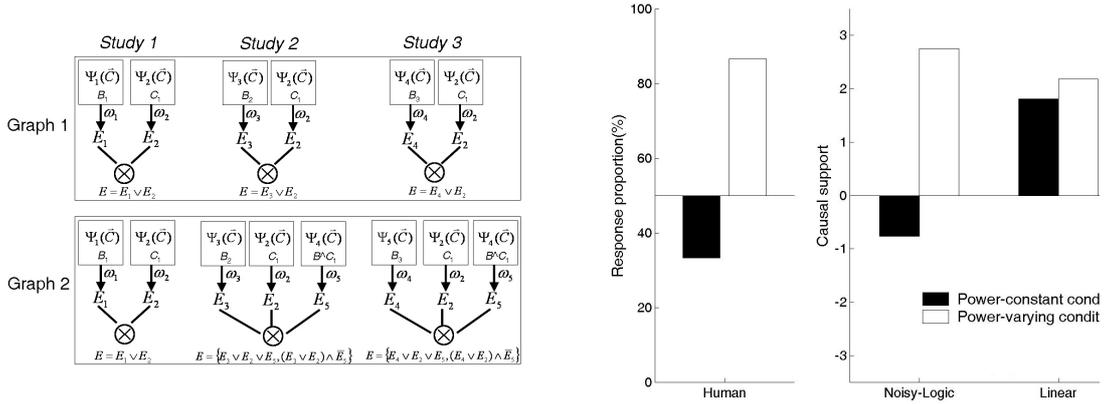


Figure 4: Model diagram and results in Experiment II. Left panel(A): Causal structure with possible causal interaction. Right panel (B): Results in a causal link judgment for interactive cause. Two experimental conditions, Power-constant condition and Power-varying condition. Left plot depicts human data (proportion of participants who responded Yes indicating the existence of causal interaction between background factors and medicine A). Right plot depicts model predictions of causal support.

6 Summary

The noisy-logical distribution gives a new way to represent conditional probability distributions defined over binary variables. The complexity of the distribution can be adjusted by restricting the set of causal factors. If all the causal factors are allowed, then the distribution can represent any conditional distribution. At the other extreme, by restricting the set of causal factors we can obtain standard distributions such as the noisy-or and noisy-and-not.

We illustrated the noisy-logical distribution by modeling experimental findings on causal reasoning. Our results showed that this distribution fitted the experimental data and, in particular, accounted for the major trends (unlike the linear model). This is consistent with the success of noisy-or and noisy-and-not models for accounting for experiments involving two causes [1], [2],[3]. This suggests that humans may make use of noisy-logical representations for causal reasoning.

One attraction of the noisy-logical representation is that it helps clarify the relationship between logic and probabilities. Standard logical relationships between causes and effects arise in the limit as the ω_i take values 0 or 1. We can, for example, bias the data towards a logical form by using a prior on the $\vec{\omega}$. This may be useful, for example, when modeling human cognition – evidence suggests that humans first learn logical relationships and, only later, move to probabilities.

In summary, the noisy-logical distribution is a novel way to represent conditional probability distributions defined on binary variables. We hope this class of distributions will be useful for modeling cognitive phenomena and for applications to artificial intelligence.

Acknowledgements

We thank Mimi Liljeholm, Patricia Cheng, Adnan Darwiche, Keith Holyoak, Iasonas Kokkinos, and YingNian Wu for helpful discussions. Mimi and Patricia kindly gave us access to their experimental data.

References

- [1] P. W. Cheng. From covariation to causation: A causal power theory. *Psychological Review*, 104, 367-405. 1997.
- [2] L.R. Novick and P.W. Cheng. Assessing interactive causal influence. *Psychological Review*, 111, 455-485. 2004.
- [3] T. L. Griffiths, and J. B. Tenenbaum. Structure and strength in causal induction. *Cognitive Psychology*, 51, 334-384, 2005.
- [4] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. Morgan-Kaufman, 1988.
- [5] M. Liljeholm and P. W. Cheng. When is a Cause the "Same"? Coherent Generalization across Contexts. *Psychological Science*, in press. 2007.
- [6] M. J. Buehner, P. W. Cheng, and D. Clifford. From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1119-1140, 2003.
- [7] P.W. Cheng. Separating causal laws from casual facts: Pressing the limits of statistical relevance. In D.L. Medin (Ed.), *The psychology of learning and motivation*, vol. 30 (pp. 215-264). New York: Academic Press. 1993.
- [8] R. A. Rescorla, and A. R. Wagner. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F. Prokasy (Eds.), *Classical conditioning II: Current theory and research* (pp. 64-99). New York: Appleton-Century Crofts. 1972.