Causal Inference and Graphical Models - II

Jin Tian

Iowa State University
Outline

- Computing the effects of manipulations
- Inferring constraints implied by DAGs with hidden variables
  - on nonexperimental data
  - on experimental data
- Determining the causes of effects
  - Counterfactuals
  - Probabilities of causation
Causal Bayesian Networks

- **Causal graph**, a DAG,
  - Nodes: random variables.
  - Edges: direct causal influence.

- **Modularity**: Each parent-child relationship represents an autonomous causal mechanism.
  - Functional: $v_i = f(pa_i, \varepsilon)$
  - Probabilistic: $P(v_i|pa_i)$
Atomic Intervention/Manipulation

\[ do(T = t) \]: fixing a set \( T \) of variables to some constants \( T = t \).

\[
P(u, x, z, y) = P(u)P(x|u)P(z|x)P(y|z, u)
\]

\[
P_{X=False}(u, z, y) = P(u)P(z|X = False)P(y|z, u)
\]
Effects of manipulations/interventions/actions

The causal effect of $T$ on $S$: $P_t(s)$.

Notations:

\[ P_t(s) = P(s|do(t)) = P(s|set(t)) = P(s|\hat{t}) = P(s||t) \]
Computing Causal Effects

Given:
- observational data: distribution \( P(v) \)
- qualitative causal assumptions: a causal graph

Can we compute the causal effect \( P_t(s) \).

Causal BNs with no hidden common causes

\[
P(v) = \prod_i P(v_i|pa_i)
\]

\[
P_t(v) = \prod_{\{i|V_i \notin T\}} P(v_i|pa_i)
\]
Computing Causal Effects

- The presence of unobserved (hidden, latent) variables.

Input: causal graph $+ P(x, y)$. Can we predict $P_x(y)$?
Computing Causal Effects

Unidentifiable

\[ P(x, y) = \sum_u P_{M1}(x|u)P_{M1}(y|x, u)P_{M1}(u) \]
\[ = \sum_u P_{M2}(x|u)P_{M2}(y|x, u)P_{M2}(u) \]
\[ P_{x_{M1}}(y) = \sum_u P_{M1}(y|x, u)P_{M1}(u) \]
\[ P_{x_{M2}}(y) = \sum_u P_{M2}(y|x, u)P_{M2}(u) \]
\[ P_{x_{M1}}(y) \neq P_{x_{M2}}(y) \]
Input: causal graph + $P(x, y, z)$. 

Identifiable – p.9
Input: causal graph + $P(x, y, z)$.
Output:

$$P_x(y) = \sum_z P(z|x) \sum_{x'} P(y|x', z) P(x')$$
Pearl’s *do*-calculus

**Rule 1: Ignoring observations**

\[ P_x(y|z, w) = P_x(y|w) \quad \text{if} \quad (Y \perp \!\!\!\!\!\!\perp Z|X, W)_{G_X} \]

**Rule 2: Action/observation exchange**

\[ P_{x,z}(y|w) = P_x(y|z, w) \quad \text{if} \quad (Y \perp \!\!\!\!\!\!\perp Z|X, W)_{G_{XZ}} \]

**Rule 3: Ignoring actions**

\[ P_{x,z}(y|w) = P_x(y|w) \quad \text{if} \quad (Y \perp \!\!\!\!\!\!\perp Z|X, W)_{G_{XZ(W)}} \]
When to use which rule of do-calculus?
For convenience of presentation, consider models in which each hidden variable is a root node and has exactly two observed children.
For convenience of presentation, consider models in which each hidden variable is a root node and has exactly two observed children.

Represent the presence of hidden variables with bidirected links.
C-components

- Variables are partitioned into c-components.
- Two variables are in the same c-components iff they are connected by a bi-directed path.
- Bi-directed path: each link on the path is a bidirected link.

Two c-components:

\[ S_1 = \{X, Z_2\} \]
\[ S_2 = \{Z_1, Y\} \]
Decomposition of $P(v)$

\[
P(v) = \sum_u \prod_{i \mid V_i \in V} P(v_i \mid pa_{v_i}) \prod_{i \mid U_i \in U} P(u_i)
\]

For any set $S \subseteq V$, define

\[
Q[S](v) = P_{v\setminus S}(s) = \sum_u \prod_{i \mid V_i \in S} P(v_i \mid pa_{v_i}) \prod_{i \mid U_i \in U} P(u_i)
\]
Decomposition of $P(v)$

$$P(v) = \sum_u \prod_{i|V_i \in V} P(v_i|pa_{v_i}) \prod_{i|U_i \in U} P(u_i)$$

For any set $S \subseteq V$, define

$$Q[S](v) = P_{v\backslash S}(s) = \sum_u \prod_{i|V_i \in S} P(v_i|pa_{v_i}) \prod_{i|U_i \in U} P(u_i)$$

**Theorem** (Decomposition of joint) Let a causal graph be partitioned into $c$-components $S_1, \ldots, S_k$. Then

$$P(v) = \prod_i Q[S_i](v) = \prod_i P_{v\backslash s_i}(s_i)$$
Decomposition of $P(\nu)$

Two c-components:

$S_1 = \{X, Z_2\}$

$S_2 = \{Z_1, Y\}$

$$P(x, y, z_1, z_2) = \sum_{u_1, u_2} P(x|u_1)P(z_1|x, u_2)P(z_2|z_1, u_1)P(y|x, z_1, z_2, u_2)P(u_1)P(u_2)$$
Decomposition of $P(v)$

Two c-components:

$S_1 = \{X, Z_2\}$

$S_2 = \{Z_1, Y\}$

\[
P(x, y, z_1, z_2) = \sum_{u_1, u_2} P(x|u_1)P(z_1|x, u_2)P(z_2|z_1, u_1) \]

\[
P(y|x, z_1, z_2, u_2)P(u_1)P(u_2)
\]

\[
= \left( \sum_{u_1} P(x|u_1)P(z_2|z_1, u_1)P(u_1) \right) \]

\[
\left( \sum_{u_2} P(z_1|x, u_2)P(y|x, z_1, z_2, u_2)P(u_2) \right)
\]

\[
= Q[S_1](x, z_1, z_2)Q[S_2](x, z_1, z_2, y)
\]

\[
= P_{y,z_1}(x, z_2)P_{x,z_2}(y, z_1)
\]
Computing $Q[S_i]$’s

**Theorem** Let a causal graph be partitioned into $c$-components $S_1, \ldots, S_k$. Then each $Q[S_i]$ is identifiable and is given by

$$Q[S_i](v) = P_{v \setminus S_i}(s_i) = \prod_{\{j|V_j \in S_i\}} P(v_j|v_1, \ldots, v_{j-1}),$$

assuming a topological order over $V$ be $V_1 < \ldots < V_n$. 
Theorem Let a topological order over $V$ be $V_1 < \ldots < V_n$,

$$P(v_i|v_1, \ldots, v_{i-1}) = P(v_i|pa(T_i) \setminus \{v_i\})$$

where $T_i$ is the c-component of the subgraph $G_{\{V_1, \ldots, V_i\}}$ that contains $V_i$.

In the presence of hidden variables, each variable is independent of its non-descendants given its parents, the non-descendant variables in its c-component, and the parents of the non-descendant variables in its c-component.
Two c-components:
\[ S_1 = \{X, Z_2\} \]
\[ S_2 = \{Z_1, Y\} \]

Topological order:
\[ X < Z_1 < Z_2 < Y \]
Two c-components:

\[ S_1 = \{X, Z_2\} \]

\[ S_2 = \{Z_1, Y\} \]

Topological order:

\[ X < Z_1 < Z_2 < Y \]

\[ P(x, y, z_1, z_2) = Q[\{X, Z_2\}]Q[\{Z_1, Y\}] \]
An Example

Two c-components:

\[ S_1 = \{X, Z_2\} \]
\[ S_2 = \{Z_1, Y\} \]

Topological order:
\[ X < Z_1 < Z_2 < Y \]

\[ P(x, y, z_1, z_2) = Q[S_X, Z_2]Q[Z_1, Y] \]

\[ Q[S_X, Z_2] = P_{y, z_1}(x, z_2) = P(x)P(z_2 | x, z_1) \]

\[ Q[Z_1, Y] = P_{x, z_2}(y, z_1) = P(z_1 | x)P(y | x, z_1, z_2) \]
Decomposition of $P_{v \setminus h}(h)$

**Theorem** Let $H \subseteq V$, and $G_H$ denote the subgraph of $G$ composed only of the variables in $H$. Assume $G_H$ is partitioned into $c$-components $H_1, \ldots, H_l$. Then

1. $Q[H] = \prod_i Q[H_i]$, i.e., $P_{v \setminus h}(h) = \prod_i P_{v \setminus h_i}(h_i)$.

2. Each $Q[H_i] = P_{v \setminus h_i}(h_i)$ is computable in terms of $Q[H] = P_{v \setminus h}(h)$. 
A procedure for computing $Q[S](\nu) = P_{\nu \backslash S}(s)$ is developed, that

1. Determine the identifiability of $Q[S]$.
2. Express identifiable $Q[S]$ in terms of $P(\nu)$. 
Let \( D = \text{An}(S)_{G_{V \setminus T}} \), and assume that the subgraph \( G_D \) is partitioned into \( c \)-components \( D_1, \ldots, D_k \). Then

\[
P_t(s) = \sum_{(v \setminus t) \setminus s} P_t(v \setminus t)
= \sum_{(v \setminus t) \setminus s} Q[V \setminus T]
= \sum_{d \setminus s} \prod_{i} Q[D_i].
\]

\( P_t(s) \) is identifiable iff each \( Q[D_i] \) is identifiable.
A complete algorithm is developed that will either determine $P_t(s)$ to be unidentifiable or express $P_t(s)$ in terms of $P(v)$.

Do-calculus is complete for computing causal effects.

Open questions:
- computing causal effects in partially known DAGs, or PAGs
Computing the effects of manipulations

Inferring constraints implied by DAGs with hidden variables

Determining the causes of effects
The validity of a causal model can be tested only if it has empirical implications, that is, it must impose constraints on data.

No hidden variables:

- observational implications of a BN are completely captured by conditional independence relationships
- read by d-separation
The validity of a causal model can be tested only if it has empirical implications, that is, it must impose constraints on data.

No hidden variables:
- observational implications of a BN are completely captured by conditional independence relationships
- read by d-separation

When hidden variables are present:
- other types of constraints on the observed distribution.
An Example

$P(a, b, c, d)$ must satisfy:

$$\sum_b P(d|a, b, c)P(b|a) = f(c, d)$$

i.e.

$$\sum_b P(d|a, b, c)P(b|a) = \sum_b P(d|a', b, c)P(b|a')$$

Functional constraints
Applications

- Empirically validating causal models.
- Distinguishing causal models with the same set of conditional independence relationships.

Independence statements: \( A \) is independent of \( C \) given \( B \).
**Inferring Functional Constraints**

Consider

\[ Q[\{D\}] = P_{a,b,c}(d) = \sum_u P(d|c,u)P(u) \equiv Q[\{D\}](c,d) \]

\[ Q[\{D\}] \] is identifiable as

\[ Q[\{D\}](v) = \sum_b P(d|a,b,c)P(b|a). \]

Therefore \( \sum_b P(d|a,b,c)P(b|a) \) is independent of \( a \).
Inferring Functional Constraints

Basic Ideas

- $Q[S](v)$ is a function of values only of a subset of $V$.

- Whenever $Q[S]$ is computable from $P(v)$, it may lead to some constraints — conditional independence relations or functional constraints.
The Arguments of $Q[S]$

$$Q[S](v) = \sum_u \prod_{\{i|V_i \in S\}} P(v_i|pa_{v_i}) \prod_{\{i|U_i \in U\}} P(u_i)$$

- $Pa(S)$: the union of $S$ and the set of parents of $S$.
- $Q[S](v)$ is a function of $Pa(S)$:

$$Q[S](v) = Q[S](pa(S))$$
Identifying Functional Constraints

1. Find a computable $Q[S]$ expressed in terms of $P(v)$
   - A procedure is developed that systematically find computable $Q[S]$.

2. $Q[S]$ is a function only of $pa(S)$
   $\iff$ conditional independence relations or functional constraints.
Another Example

The model does not imply any conditional independences

\[ Q[\{V_4\}](v_3, v_4) = \frac{\sum_{v_1} P(v_4|v_3, v_2, v_1)P(v_3|v_2, v_1)P(v_1)}{\sum_{v_1} P(v_3|v_2, v_1)P(v_1)}. \]

The right hand side is independent of \( v_2 \).
Pearl’s *instrumental inequality*, for discrete variables

\[
\max_x \sum_y \max_z P(xy|z) \leq 1.
\]

E.g., binary variables

\[
\begin{align*}
P(x_0, y_0|z_0) + P(x_0, y_1|z_1) &\leq 1 \\
P(x_1, y_0|z_0) + P(x_1, y_1|z_1) &\leq 1 \\
P(x_0, y_1|z_0) + P(x_0, y_0|z_1) &\leq 1 \\
P(x_1, y_1|z_0) + P(x_1, y_0|z_1) &\leq 1
\end{align*}
\]
Inequality Constraints

- Empirically validating causal models.
- Distinguishing causal models with the same set of conditional independence relationships.
- Open problem: how to identify inequality constraints
A causal BN not only imposes constraints on the nonexperimental distribution but also on the experimental distributions.

A causal BN can be tested and falsified by using two types of data:
- nonexperimental data are passively observed,
- experimental data are produced by manipulating (randomly) some variables and observing the states of other variables.

The ability to use a mixture of nonexperimental and experimental data will greatly increase our power of causal reasoning and learning.
Constraints on Experimental Data

Let $H \subseteq V$ and assume the subgraph $G_H$ is partitioned into $c$-components $H_1, \ldots, H_l$. Then

$$P_{v \setminus h}(h) = \prod_i P_{v \setminus h_i}(h_i).$$

$$P_{pa_i,s}(v_i) = P_{pa_i}(v_i), \forall S \subseteq V \setminus (PA_i \cup \{V_i\})$$

If a set $T$ is composed of non-descendants of $V_j$, then

$$P_{v_j,s}(t) = P_s(t).$$
Constraints on Experimental Data

\[ P_z(xy) = P(xy|z) \]
\[ P_{yz}(x) = P(x|z) \]
\[ P_{xz}(y) = P_x(y) \]
Inequalities on Experimental Data

Consider discrete random variables

A type of inequality constraints on experimental distributions

Let \( V \) be partitioned into \( c \)-components \( T_1, \ldots, T_k \). For \( i = 1, \ldots, k \), \( \forall S_1 \subseteq T_i \),

\[
\sum_{S_2 \subseteq T_i \setminus S_1} (-1)^{|S_2|} P_{v \setminus (s_1 \cup s_2)}(s_1, s_2) \geq 0, \quad \forall v \in Dm(V)
\]

Not complete
For all $x \in Dm(X)$, $y \in Dm(Y)$, $z \in Dm(Z)$

\[1 - P_{yz}(x) - P_{xz}(y) + P_z(xy) \geq 0\]

\[P_{yz}(x) - P_z(xy) \geq 0\]

\[P_{xz}(y) - P_z(xy) \geq 0\]
Applications of Inequalities

- Model testing using a mixture of nonexperimental and experimental data
- Bounding (unidentifiable) causal effects from nonexperimental data
- Bounding the effects of untried interventions from experiments involving auxiliary interventions that are easier or cheaper to implement

\[ P_z(x, y) \leq P_{xz}(y) \leq 1 - P_z(x) + P_z(x, y) \]
Deriving Instrumental Inequality

Equality constraints: $P_z(xy) = P(xy|z)$, $P_{xz}(y) = P_x(y)$

Inequality: $P_z(xy) \leq P_{xz}(y)$

We have

\[
\begin{align*}
P(xy|z) & \leq P_x(y) \\
\max_z P(xy|z) & \leq P_x(y) \\
\sum_y \max_z P(xy|z) & \leq 1
\end{align*}
\]
The following instrumental type inequality can be derived

$$
\sum_{yz} \max_{w_1} P(z|w_1xw_2y) P(y|w_1xw_2) P(x|w_1) \leq 1.
$$
Experimental Implications

- What if causal structures unknown?
- Given a collection of experimental distributions
  \[ P_* = \{P_t(v) | T \subseteq V, t \in Dm(T)\} \]
- Is the collection \(P_*\) compatible with some underlying causal Bayesian network?
Three Properties

- If no hidden variables

1. Effectiveness

   \( P_t(t) = 1. \)

2. Markov

   \[ P_{v\setminus(s_1 \cup s_2)}(s_1, s_2) = P_{v\setminus s_1}(s_1) P_{v\setminus s_2}(s_2) \]

3. Recursiveness

   Define \( X \leadsto Y \) as \( \exists w, P_{x,w}(y) \neq P_w(y), \)

   \[ (X_0 \leadsto X_1) \land \ldots \land (X_{k-1} \leadsto X_k) \Rightarrow \lnot(X_k \leadsto X_0) \]
Theorem (Soundness) Effectiveness, Markov, and recursiveness hold in all causal Bayesian networks.

Theorem (Completeness) If a \( P_\star \) set satisfies effectiveness, Markov, and recursiveness, then there exists a causal Bayesian network with a unique causal graph that can generate this \( P_\star \) set.
Effectiveness

Recursiveness

Directionality

There exists a total order "<" such that

\[ P_{v_i,w}(s) = P_w(s) \quad \text{if} \quad \forall X \in S, X < V_i, \]

Inclusion-Exclusion Inequalities

For any subset \( S_1 \subseteq V \),

\[
\sum_{S_2 \subseteq V \setminus S_1} (-1)^{|S_2|} P_{v \setminus (s_1 \cup s_2)}(v) \geq 0, \quad \forall v \in Dm(V),
\]
Theorem (Soundness) Effectiveness, recursiveness, directionality, and inclusion-exclusion inequalities hold in all semi-Markovian models.

Theorem (Completeness) If a $P_*$ set satisfies effectiveness, recursiveness, directionality, and inclusion-exclusion inequalities, then there exists a semi-Markovian model that can generate this $P_*$ set.
Applications of Characterization

- Reasoning about causal effects without possessing causal structures
- Is a collection of experimental distributions compatible?
- Predicting about or bounding interventions that were not tried experimentally even if the structure of the underlying model is unknown
Open Problems

- Identifying *all* constraints
  - on nonexperimental distributions
  - on experimental distributions
  - equalities
  - inequalities
  - constraints particular to a family of distributions

- Using constraints to guide learning BNs with hidden variables
Computing the effects of manipulations
Inferring constraints implied by DAGs with hidden variables
Determining the causes of effects
  Counterfactuals
  Probabilities of causation
Determining the Causes of Effects

- Assessing the likelihood that one event was the cause of another
- Legal responsibility: Mr. A took a drug and died,
  - Lawsuit: the drug caused the death of Mr. A
  - Experimental and nonexperimental data on patients
- Court to decide:
  Is it more probable than not that A would be alive but for the drug?
The Problem

Probability of necessary causation (PN):
“Probability that event $y$ would not have occurred if it were not for event $x$, given that $x$ and $y$ did in fact occur.”

What is the meaning of $PN$? How to define PN mathematically?

Under what conditions can $PN$ be learned from statistical data?
Functional Causal Models

- Structural Equations
  \[ v_i = f_i(pa_i, u_i), \quad i = 1, \ldots, n. \]
- \( U = \{U_1, \ldots, U_n\} \): exogenous background/error variables
- Acyclic models
  - The values of the \( V \) variables will be uniquely determined by those of the \( U \) variables.
  - The joint distribution \( P(v) \) is determined uniquely by the distribution \( P(u) \).
  - \( P(u) \) defines a probabilistic causal model
An intervention is represented as an alteration on a select set of functions instead of a select set of conditional probabilities.

The effect of \( do(V_i = v_i) \) is represented by replacing the equation \( v_i = f_i(pai, ui) \) with

\[
V_i = v_i
\]

The counterfactual expression “The value that \( Y \) would have obtained, had \( X \) been \( x \”), denoted by \( Y_x(u) \), is interpreted as the solution for \( Y \) in the modified set of equations in situation \( U = u \).
Probabilities of Counterfactuals

\[ P(Y = y) = \sum_{\{u \mid Y(u) = y\}} P(u) \]

\[ P(Y_x = y) = \sum_{\{u \mid Y_x(u) = y\}} P(u) \equiv P_x(y) \]

\[ P(Y_x = y, X = x') = \sum_{\{u \mid Y_x(u) = y \& X(u) = x'\}} P(u) \]

\[ P(Y_x = y, Y_{x'} = y') = \sum_{\{u \mid Y_x(u) = y \& Y_{x'}(u) = y'\}} P(u) \]
Computing Counterfactuals

Given evidence $X = x', Y = y'$, compute the probability of $Y = y$ had $X$ been $x$ ($X$ and $Y$ subsets of variables):

**Step 1** (*abduction*): Update the probability $P(u)$ to obtain $P(u|x', y')$.

**Step 2** (*action*): Replace the equations corresponding to variables in set $X$ by the equations $X = x$.

**Step 3** (*prediction*): Use the modified model to compute the probability of $Y = y$. 
Computing Counterfactuals

Model 1

\[ x = u_1, \]
\[ y = u_2. \]

Model 2

\[ x = u_1, \]
\[ y = xu_2 + (1 - x)(1 - u_2). \]

where \( U_1 \) and \( U_2 \) are two independent binary variables with \( P(u_1 = 1) = P(u_2 = 1) = \frac{1}{2} \), leading to the same distribution \( P(x, y) \).

Model 1: \( P(Y_{x=0} = 0|X = 1, Y = 1) = 0 \)

Model 2: \( P(Y_{x=0} = 0|X = 1, Y = 1) = 1 \)
Probabilistic causal models are insufficient for computing probabilities of counterfactuals; knowledge of the actual process behind $P(y|x)$ is needed for the computation.

A functional causal model constitutes a mathematical object sufficient for the computation and definition of such probabilities.
Let $X$ and $Y$ be two binary variables

**Probability of necessity (PN)**

$$PN \equiv P(Y_{x'} = y' \mid X = x, Y = y) \equiv P(y'_{x'} \mid x, y)$$

PN stands for the probability that event $y$ would not have occurred in the absence of event $x$, $y'_{x'}$, given that $x$ and $y$ did in fact occur.

Applications in epidemiology, legal reasoning, and AI: a certain case of disease is *attributable* to a particular exposure, “the probability that disease would not have occurred in the absence of exposure, given that disease and exposure did in fact occur.”
Probabilities of Causation

- Probability of sufficiency (PS)

\[ PS \equiv P(y_x|y', x') \]

PS gives the probability that setting \( x \) would produce \( y \) in a situation where \( x \) and \( y \) are in fact absent.

Applications in policy analysis, AI, and psychology: a policy maker interested in the dangers that a certain exposure may present to the healthy population, the “probability that a healthy unexposed individual would have gotten the disease had he/she been exposed.”
A lawsuit is filed against the manufacturer of drug $x$, charging that the drug is likely to have caused the death of Mr. A, who took the drug to relieve symptom $S$ associated with disease $D$.

Experimental and nonexperimental data (in the next page)

Court to decide:
Is it more probable than not that A would be alive but for the drug?

Can PN be estimated from data?
Table 0: (Hypothetical) frequency data obtained in experimental and nonexperimental studies, comparing deaths (in thousands) among drug users, $x$, and non-users, $x'$.

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Nonexperimental</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x$</td>
<td>$x'$</td>
</tr>
<tr>
<td>Deaths ($y$)</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Survivals ($y'$)</td>
<td>984</td>
<td>986</td>
</tr>
</tbody>
</table>
Parameters: \( p_{110} = P(y_x, y_{x'}, x') \), \( \ldots \)

Probabilistic constraints:

\[
\sum_{i=0}^{1} \sum_{j=0}^{1} \sum_{k=0}^{1} p_{ijk} = 1
\]

\[
p_{ijk} \geq 0 \text{ for } i, j, k \in \{0, 1\}
\]

Nonexperimental constraints:

\[
p_{111} + p_{101} = P(x, y)
\]

\[
p_{011} + p_{001} = P(x, y')
\]

\[
p_{110} + p_{010} = P(x', y)
\]
Bounding by LP

- Experimental constraints:

\[
P(y_x) = p_{111} + p_{110} + p_{101} + p_{100}
\]
\[
P(y_{x'}) = p_{111} + p_{110} + p_{011} + p_{010}
\]

- Maximize (Minimize)

\[
PN = p_{101}/P(x, y)
\]
\[
PS = p_{100}/P(x', y')
\]
Bounds on the probabilities of causation given combined nonexperimental and experimental data

\[
\max \left\{ \frac{0}{P(y) - P(y_x')} \right\} \leq PN \leq \min \left\{ \frac{1}{P(y'_x) - P(x', y')} \right\}
\]

\[
\max \left\{ \frac{0}{P(y_x) - P(y)} \right\} \leq PS \leq \min \left\{ \frac{1}{P(y_x) - P(x, y)} \right\}
\]
Solution to Legal Responsibility

- **Plaintiff:**

\[
P_N \geq \frac{P(y) - P(y_{x'})}{P(y, x)} = \frac{0.015 - 0.014}{0.001} = 1
\]

- **Jury:** Guilty!
Mr. \( B \), survived without drug. Would he risk death by starting now?

- Nonexperimental data: \( P(y|x) = 0.002 \)
- Experimental data: \( P(y_x) = 0.016 \)
- Correct Answer: Risk = \( PS = P(y_x|x', y') \)

\[
0.002 \leq PS \leq 0.031
\]
**Hierarchy of Causal Queries**

- **Predictions** (conditioning) require only a specification of a joint distribution function.

- **Intervention** analysis requires a causal structure in addition to a joint distribution.

- **Counterfactual** analysis requires information about the functional relationships and the distribution of the omitted factors.