Causal Inference in the Decision-Theoretic Framework of Statistical Causality

Nayia Constantinou

University of Bristol

19th December 2014
Probability and statistics seminars

Outline

Correlation vs Causation

"Correlation does not imply Causation": For an observed association between two variables of interest (e.g. a risk factor X and a disease status Y), it is possible to argue in totally different directions:

- X is causal for Y (causation)
- Y is causal for X (reverse causation)
- U is causal for both X and Y (confounding)

Detection and assessment of causality is fundamental for the design of public health intervention policies. Causal inference is a central aim for most medical and epidemiological studies.

For the purposes of this talk, we consider as causal the effect of an intervention.

Intervention and Confounding variables

Intervention: Controlling some quantity to take on a particular value (as oppose to observing it taking this value naturally).

Example (Lindley, 2002): Let X denote the size of foot and Y denote the size of hand. The binding of feet, practiced in some societies to lessen X, has no effect on size of hand. Thus

- $p(y \mid x) \neq p(y \mid do(X = x))$
- But $p(y \mid do(X = x))$ could be $p(y)$

Pearl (1997) writes $p(y \mid do(X = x))$ when we force X to be equal to x and $p(y \mid see(X = x))$ when we see X being equal to x.

Rubin (1974) introduces the Potential Responses Framework, where $\mathbf{Y} = (Y_0, Y_1)$.

Aiming to detect, assess and compare the effects of different interventions, extracting "causal conclusions" from observational data might be seriously misleading!

Observational data and confounding

Observational data: Data that have been gathered under conditions which the statistician was not able to control.

Examples:
- Doctor assigns treatment only to patients that look healthier.
- Treatment is taken only by patients who wish to take it.

Problem: In such cases we will not know whether to ascribe observed differences between the responses in different treatment groups to the treatments themselves, or to other differences between the groups, whose effects would persist even if all units were treated identically.

Confounding: Variables that we have failed to account or control that may damage the validity of our "causal" conclusions.

The Decision-Theoretic Framework

Dawid (2000, 2007a)

Example:

$$T = \begin{cases} 0, & \text{control treatment} \\ 1, & \text{active treatment} \end{cases} \qquad Y = \begin{cases} 0, & \text{no illness} \\ 1, & \text{illness} \end{cases}$$

Question: Will intervention on the treatment variable T cause an effect on the illness outcome Y?

Aim: Compare the distribution of Y in the two interventional regimes.

The Decision-Theoretic Framework

Consider an index variable Σ to differentiate between the different regimes.

$$\Sigma = \begin{cases} \emptyset, & \text{observational regime.} \\ 0, & \text{interventional regime under control treatment.} \\ 1, & \text{interventional regime under active treatment.} \end{cases}$$

Note: Any probabilistic statement about the stochastic variables of interest must be conditional on the regime.

We want to explore under which conditions we can extract information from the observational regime for the interventional regimes.

Average Causal Effect (1/3)

Identification of the Average Causal Effect (ACE), defined by:

$$ACE := \mathbb{E}_1(Y) - \mathbb{E}_0(Y)$$
$$= \mathbb{E}_1(Y \mid T = 1) - \mathbb{E}_0(Y \mid T = 0) \quad (\text{since } \mathbb{P}_t(T = t) = 1)$$

Assumption: Given treatment T, Y has the same distribution in observational and interventional regimes. Notation: $Y \perp\!\!\!\perp \Sigma \mid T$.

Then

$$ACE = \mathbb{E}_\emptyset(Y \mid T = 1) - \mathbb{E}_\emptyset(Y \mid T = 0)$$

and we have achieved identification of a causal quantity from observational data.

Average Causal Effect (1/3)

Identification of the Average Causal Effect (ACE), defined by:

$$ACE := \mathbb{E}_1(Y) - \mathbb{E}_0(Y)$$
$$= \mathbb{E}_1(Y \mid T = 1) - \mathbb{E}_0(Y \mid T = 0) \quad (\text{since } \mathbb{P}_t(T = t) = 1)$$

Assumption: Given treatment T, Y has the same distribution in observational and interventional regimes. Notation: $Y \perp\!\!\!\perp \Sigma \mid T$.

Then

$$ACE = \mathbb{E}_\emptyset(Y \mid T = 1) - \mathbb{E}_\emptyset(Y \mid T = 0)$$

and we have achieved identification of a causal quantity from observational data.

Average Causal Effect (2/3)

Problem: In most contexts $Y \perp\!\!\!\perp \Sigma \mid T$ is an assumption which don't believe and thus are not in position to defend. It would be valid if the observational regime represented a Randomized Controlled Trial (where patients are randomly assigned to treatment) but in general this would not be the case.

Way out: We can explore different conditional independence properties which we believe better represent the problem under study and explore if under those conditions we can still identify the ACE from observational data. For example, we can consider Strongly Sufficient Covariates (Guo and Dawid, 2010).

Average Causal Effect (3/3)

Strongly Sufficient Covariates: Additional variable(s) U that are present in the observational regime and can influence the choice for treatment, as well as, the response variable we are interested in. In particular, they have the following properties:

i) $U \perp\!\!\!\perp \Sigma$

ii) $Y \perp\!\!\!\perp \Sigma \mid (U, T)$

iii) For $t = 0, 1$, $P_\emptyset(T = t \mid U) > 0$ a.s. $[\mathbb{P}_\emptyset]$

Assuming the existence of a strongly sufficient covariate we can achieve identification of the ACE:

$$ACE = \mathbb{E}_1(Y) - \mathbb{E}_0(Y)$$

$$\vdots$$

$$= \mathbb{E}_\emptyset(Y \mid T = 1) - \mathbb{E}_\emptyset(Y \mid T = 0)$$

Conditional Independence

- ▶ Conditional independence for stochastic variables has been widely studied in probability theory and the properties that accrue from this notion are well-known.

- ▶ In the DT-Framework we allow stochastic and non-stochastic variables together. Aim: Formally extend the language to incorporate stochastic and non-stochastic variables simultaneously and explore if under the extended setting the same calculus accrues (Dawid, 1979, 1980, 2004).

- ▶ Sufficiency: Let $\boldsymbol{X} := X_1, X_2, \ldots, X_n$ be a random sample from a probability distribution with unknown parameter θ and consider having reduced information through a statistic $T = T(\boldsymbol{X})$. We say that T is sufficient for θ if:

$$P_\theta(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = T(\boldsymbol{x})) = P(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = T(\boldsymbol{x}))$$

Conditional Independence

- Conditional independence for stochastic variables has been widely studied in probability theory and the properties that accrue from this notion are well-known.

- In the DT-Framework we allow stochastic and non-stochastic variables together. Aim: Formally extend the language to incorporate stochastic and non-stochastic variables simultaneously and explore if under the extended setting the same calculus accrues (Dawid, 1979, 1980, 2004).

- Sufficiency: Let $\boldsymbol{X} := X_1, X_2, \ldots, X_n$ be a random sample from a probability distribution with unknown parameter θ and consider having reduced information through a statistic $T = T(\boldsymbol{X})$. We say that T is sufficient for θ if:

$$P_\theta(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = T(\boldsymbol{x})) = P(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = T(\boldsymbol{x}))$$

Stochastic Conditional Independence

Definition
Let X, Y, Z be random variables on $(\Omega, \mathcal{A}, \mathbb{P})$. We say that X is (conditionally) independent of Y given Z and write $X \perp\!\!\!\perp Y \mid Z$ if for all (real, bounded and measurable) functions $f(X)$,

$$\mathbb{E}(f(X) \mid Y, Z) = \mathbb{E}(f(X) \mid Z) \quad \text{a.s..}$$

Definition (Discrete Case)
$X \perp\!\!\!\perp Y \mid Z$ if for all events $\{X = x\}$,

$$P(X = x \mid Y = y, Z = z) = P(X = x \mid Z = z)$$

whenever $P(Y = y, Z = z) > 0$.

Axioms of Conditional Independence

Dawid (1979, 1980)

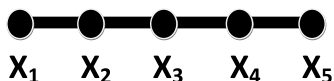Notation: $W \preceq Y$ means that there exists f such that $W = f(Y)$.

Theorem
Let X, Y, Z and W be random variables. Then the following properties hold.

1. *(P1-Symmetry) $X \perp\!\!\!\perp Y \mid Z \Rightarrow Y \perp\!\!\!\perp X \mid Z$.*
2. *(P2) $X \perp\!\!\!\perp Y \mid X$.*
3. *(P3-Decomposition) $X \perp\!\!\!\perp Y \mid Z$, $W \preceq Y \Rightarrow X \perp\!\!\!\perp W \mid Z$.*
4. *(P4-Weak Union) $X \perp\!\!\!\perp Y \mid Z$, $W \preceq Y \Rightarrow X \perp\!\!\!\perp Y \mid (W, Z)$.*
5. *(P5-Contraction) $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp W \mid (Y, Z) \Rightarrow X \perp\!\!\!\perp (Y, W) \mid Z$.*

Example

Nearest Neighbour Property of a Markov Chain: Let X_1, X_2, X_3, X_4, X_5 be random variables and suppose that:

$$\mathbf{X_1} \quad \mathbf{X_2} \quad \mathbf{X_3} \quad \mathbf{X_4} \quad \mathbf{X_5}$$

i) $X_3 \perp\!\!\!\perp X_1 \mid X_2$,

ii) $X_4 \perp\!\!\!\perp (X_1, X_2) \mid X_3$,

iii) $X_5 \perp\!\!\!\perp (X_1, X_2, X_3) \mid X_4$.

Then we can show that $X_3 \perp\!\!\!\perp (X_1, X_5) \mid (X_2, X_4)$, just by using the axioms.

Extended Conditional Independence: Notation

- Regime space $\mathcal{S} = \{\sigma_i : i \in I\}$, where I is a set of indices (e.g. observational/interventional regimes).

- Non-stochastic variables which are functions defined on $\mathcal{S}$ (say $\Theta : \mathcal{S} \to \Theta(\mathcal{S})$) which are called *decision variables* (e.g. regime indicator Σ).

- Stochastic variables defined on $(\Omega, \mathcal{A})$, which have different distributions under the different regimes $\sigma \in \mathcal{S}$ (e.g. treatment variable T, disease indicator Y).

- Family of $\mathbb{P}$-measures defined on $(\Omega, \mathcal{A})$ indexed by $\sigma \in \mathcal{S}$ and denoted by $\mathbb{P}_\sigma$.

Extended Conditional Independence

Definition

Let X, Y and Z be stochastic variables and let Σ be the regime indicator. We say that X is (conditionally) independent of (Y, Σ) given Z and write $X \perp\!\!\!\perp (Y, \Sigma) \mid Z$ if for all (real, bounded and measurable) functions $h(X)$, there exists a function $w(Z)$ such that for all $\sigma \in \mathcal{S}$,

$$\mathbb{E}_\sigma[h(X) \mid Y, Z] = w(Z) \quad \text{a.s.} \ [\mathbb{P}_\sigma]$$

Note: In fact, we can further generalise this definition to $X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$.

Validity of the axioms: Discrete regime space

Bayesian argument:

We consider further to to $(\Omega, \mathcal{A})$ and $\mathcal{S}$, $\mathcal{F} := \sigma(\mathcal{S})$ and we define an arbitrary $\mathbb{P}$-measure on $\mathcal{F}$. Then we consider the product space $\Omega \times \mathcal{S}$ with its corresponding σ-algebra $\mathcal{A} \otimes \mathcal{F}$ (where $\mathcal{A} \otimes \mathcal{F} := \sigma(\mathcal{A} \times \mathcal{F}) := \sigma(\{A \times B : A \in \mathcal{A}, B \in \mathcal{F}\})$). Extending the space in this way allows us to consider both the stochastic variables $X, Y, Z, \ldots$ and the decision variables $\Theta, \Phi, \ldots$ as measurable functions defined on $(\Omega \times \mathcal{S}, \mathcal{A} \otimes \mathcal{F})$. Let us denote them by $X^*, Y^*, Z^*, \ldots$ and $\Theta^*, \Phi^*, \ldots$. We further consider the corresponding $\mathbb{P}^*$-measure for $(\Omega \times \mathcal{S}, \mathcal{A} \otimes \mathcal{F})$.

Theorem
$X \perp\!\!\!\perp (Y, \Theta) \mid (Z, \Phi)$ *(ECI) if and only if* $X^* \perp\!\!\!\perp (Y^*, \Theta^*) \mid (Z^*, \Phi^*)$ *(SCI)*.

Note: The assumption of a discrete regime space is crucial.

Validity of the axioms

Theorem

Suppose that one of the following conditions holds:

- ▶ *the regime space $\mathcal{S}$ is discrete*
- ▶ *the random variables are discrete*
- ▶ *there exists a dominating regime*

Further suppose we are given a collection of extended conditional independence properties. Any deduction made using the axioms of stochastic conditional independence will be valid, so long as, in both premisses and conclusions, no non-stochastic variables appear in the left-most term in a conditional independence statement.

Sequential decision problems

Robins (1986, 1987, 1992); Murphy (2003); Dawid and Didelez (2010)

Sequence of domain variables: $L_1, A_1, \ldots, L_n, A_n, L_{n+1} \equiv Y$

- ▶ Observable variables $(L_1, L_2, \ldots)$
- ▶ Action variables $(A_1, A_2, \ldots)$

Aim: Evaluate the consequence of a strategy on the variable of interest Y.

Regime indicator Σ, values in $\mathcal{S} = \{\emptyset\} \cup \mathcal{S}^*$

- ▶ $\emptyset$ is the observational regime under which data have been gathered
- ▶ $\mathcal{S}^*$ is a collection of interventional strategies of interest.

We would like to use data gathered under $\emptyset$ to infer what would happen if a strategy $e \in \mathcal{S}^*$ were applied.

Sequential decision problems

Robins (1986, 1987, 1992); Murphy (2003); Dawid and Didelez (2010)

Sequence of domain variables: $L_1, A_1, \ldots, L_n, A_n, L_{n+1} \equiv Y$

- ▶ Observable variables $(L_1, L_2, \ldots)$
- ▶ Action variables $(A_1, A_2, \ldots)$

Aim: Evaluate the consequence of a strategy on the variable of interest Y.

Regime indicator Σ, values in $\mathcal{S} = \{\emptyset\} \cup \mathcal{S}^*$

- ▶ $\emptyset$ is the observational regime under which data have been gathered
- ▶ $\mathcal{S}^*$ is a collection of interventional strategies of interest.

We would like to use data gathered under $\emptyset$ to infer what would happen if a strategy $e \in \mathcal{S}^*$ were applied.

Identification of an interventional strategy

Suppose we are interested in computing $\mathrm{E}_e(Y)$ for some interventional strategy $e \in \mathcal{S}^*$.

We need $p_e(y)$ which can be computed from $p_e(y, \overline{l}_n, \overline{a}_n)$, where $\overline{l}_i := l_1, \ldots, l_i$ and $\overline{a}_i := a_1, \ldots, a_i$.

The joint density can be factorised as follows:

$$p_e(y, \overline{l}_n, \overline{a}_n) = \left\{ \prod_{i=1}^{n+1} p_e(l_i \mid \overline{l}_{i-1}, \overline{a}_{i-1}) \right\} \times \left\{ \prod_{i=1}^{n} p_e(a_i \mid \overline{l}_i, \overline{a}_{i-1}) \right\}.$$

Identification of an interventional strategy

Suppose we are interested in computing $E_e(Y)$ for some interventional strategy $e \in \mathcal{S}^*$.

We need $p_e(y)$ which can be computed from $p_e(y, \overline{l}_n, \overline{a}_n)$, where $\overline{l}_i := l_1, \ldots, l_i$ and $\overline{a}_i := a_1, \ldots, a_i$.

The joint density can be factorised as follows:

$$p_e(y, \overline{l}_n, \overline{a}_n) = \left\{ \prod_{i=1}^{n+1} p_e(l_i \mid \overline{l}_{i-1}, \overline{a}_{i-1}) \right\} \times \left\{ \prod_{i=1}^{n} p_e(a_i \mid \overline{l}_i, \overline{a}_{i-1}) \right\}.$$

G-recursion (1/2)

An alternative way to compute $\mathrm{E}_e(Y)$ is by using "*G*-recursion", the backwards induction routine of dynamic programming (Robins, 1986, 1992):

- Denote by h a partial history of the form $(\bar{l}_i, \bar{a}_{i-1})$ or $(\bar{l}_i, \bar{a}_i)$.

- Define $f(h) := \mathrm{E}_e(Y \mid h)$.

- For h a full history $(\bar{l}_{n+1}, \bar{a}_n)$, we have that $f(h) = y$.

- For h the empty history, we have that $f(\emptyset) = \mathrm{E}_e(Y)$.

G-recursion (2/2)

- Applying the laws of probability, we get:

$$f(\bar{l}_i, \bar{a}_{i-1}) = \sum_{a_i} p_e(a_i \mid \bar{l}_i, \bar{a}_{i-1}) \times f(\bar{l}_i, \bar{a}_i) \qquad (1)$$

$$f(\bar{l}_{i-1}, \bar{a}_{i-1}) = \sum_{l_i} p_e(l_i \mid \bar{l}_{i-1}, \bar{a}_{i-1}) \times f(\bar{l}_i, \bar{a}_{i-1}) \quad (2)$$

- Use $f(\bar{l}_{n+1}, \bar{a}_n)$ as starting values.

- Successively implement (1) and (2) in turn, starting with (2) for $i = n+1$ and ending with (2) for $i = 1$.

- We exit with $f(\emptyset) = \mathrm{E}_e(Y)$.

Identification of $\mathrm{E}_e(Y)$

In order to compute $\mathrm{E}_e(Y)$ directly or using G-recursion, we need:

   i) $p_e(a_i \mid \overline{l}_i, \overline{a}_{i-1})$ for $i = 1, \ldots, n$

  ii) $p_e(l_i \mid \overline{l}_{i-1}, \overline{a}_{i-1})$ for $i = 1, \ldots, n+1$

- $p_e(a_i \mid \overline{l}_i, \overline{a}_{i-1})$: specified by the strategy

- $p_e(l_i \mid \overline{l}_{i-1}, \overline{a}_{i-1})$: unknown and in the absence of interventional data that represent the strategy, we want to explore conditions that will allow us to compute it from the observational regime.

Simple stability

The problem exhibits (simple) stability if, with Σ denoting the (non-random) regime indicator, taking values in $\mathcal{S} = \{o\} \cup \mathcal{S}^*$,

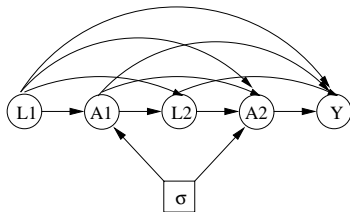$$L_i \perp\!\!\!\perp \Sigma \mid (\overline{L}_{i-1}, \overline{A}_{i-1}) \quad (i = 1, \ldots, n+1).$$

Figure : Stability

Does not take into account unobserved variables denoted by U_i $(i = 1, \ldots, N)$.

Extended stability

▶ Stability with respect to partially unobserved information:

$$(U_i, L_i) \perp\!\!\!\perp \Sigma \mid (\overline{U}_{i-1}, \overline{L}_{i-1}, \overline{A}_{i-1}).$$

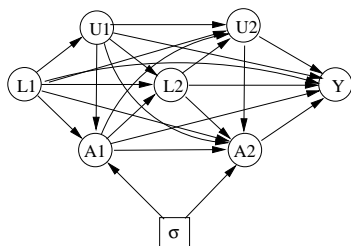▶ Does not imply simple stability.

Figure : Extended stability

Stability regained. Sequential randomization

$$A_i \perp\!\!\!\perp \overline{U}_i \mid (\overline{L}_i, \overline{A}_{i-1} ; \Sigma) \quad (i = 1, \ldots, N).$$
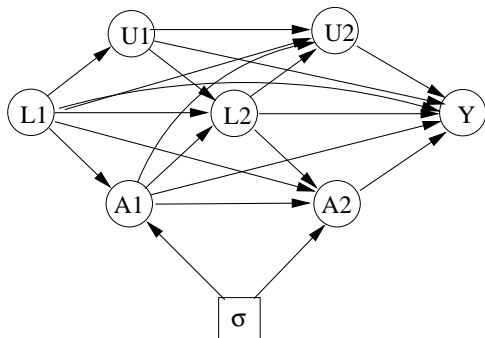

Figure : Sequential randomization

► Extended stability and sequential randomization imply simple stability (proof by induction).

Stability regained? Sequential irrelevance

$$L_i \perp\!\!\!\perp \overline{U}_{i-1} \mid (\overline{L}_{i-1}, \overline{A}_{i-1} \,;\, \Sigma) \quad (i = 1, \ldots, n+1).$$
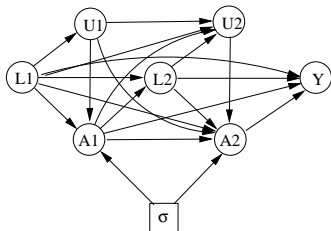

Figure : Sequential irrelevance

▶ If random variables are discrete, we can rigorously deduce simple stability from extended stability and sequential irrelevance without requiring additional assumptions. If not, we need to bring in additional assumptions (e.g. absolute continuity) which involve the unobservable variables (Dawid and Constantinou, 2014).

The Regression Discontinuity Design

Work with Aidan O'Keeffe.

- The "Regression Discontinuity Design" (RD Design) is a design that specifies a threshold above or below which an intervention is assigned. By comparing observations lying closely on either side of the threshold, it is possible to make causal inference.

- The RD Design was first introduced as a method in econometrics during the 1960s. Original idea: "exploit policy thresholds to estimate the causal effect of a particular intervention".

- Medicines might be prescribed according to pre-defined rules/guidelines (possibly government-defined). For example:
  - Antiretroviral HIV drugs might be prescribed when a patient's CD4 count is less than 200 cells/mm3.
  - Statins might be prescribed when a patient's 10-year risk of a cardiovascular event (10-year CVD risk score) exceeds 20 %.

Sharp and fuzzy design

Sharp Design:

- ► Threshold behaves like a randomising device.
- ► If the threshold is adhered to very strictly (sharp design), then we can think of the RD design as removing the confounding due to unobserved factors.

Fuzzy Design:

- ► Threshold doesn't behave like a randomising device.
- ► In medicine, sharp threshold is unlikely to be adhered to. For example, often GPs override guidelines (generally because, contrary to their recommendations, they feel that patients will benefit from medication).
- ► Statins example: some patients might request statins irrespective of their 10-year CVD risk score.

Summary

- The need to express causal concepts necessitates the use of an appropriate framework which differentiates between seemingly related observations and causally related observations.

- The DT framework makes this distinction by introducing a non-stochastic variable to index the differing regimes (observational and interventional).

- In practice, not usually being able to obtain data from the interventional regimes of interest, we want to explore under which conditions we can deduce information for the interventional regimes using the observational regime.

- We can formally express and explore these conditions using the language and calculus of conditional independence.

Summary

Using this language we can discuss:

- Sequential Decision Problems where we are concerned with controlling a variable of interest via a multi-stage procedure.

    - We have explored conditions called stability, extended stability, sequential randomisation and sequential irrelevance.

- Strict and fuzzy Regression Discontinuity Designs where we are concerned with exploiting policy thresholds to estimate the causal effect of a particular intervention.

THANK YOU!

Bibliography I

Constantinou, P. (2013). Conditional Independence and Applications in Statistical Causalily. PhD thesis, University of Cambridge.

Dawid, A. and Constantinou, P. (2014). A formal treatment of sequential ignorability. Statistics in Biosciences, 6(2):166–188.

Dawid, A. P. (1979). Conditional independence in statistical theory. Journal of the Royal Statistical Society. Series B (Methodological), 41(1):1–31.

Dawid, A. P. (1980). Conditional independence for statistical operations. The Annals of Statistics, 8(3):598–617.

Dawid, A. P. (2000). Causal inference without counterfactuals (with discussion). Journal of the American Statistical Association, 95:407 – 424.

Bibliography II

Dawid, A. P. (2001a). Separoids: A mathematical framework for conditional independence and irrelevance. Annals of Mathematics and Artificial Intelligence, 32:335–372.

Dawid, A. P. (2001b). Some variations on variation independence. In Proceedings of Artificial Intelligence and Statistics 2001, pages pp. 187 – 191. Morgan Kaufmann Publishers.

Dawid, A. P. (2004). Conditional Independence. John Wiley & Sons, Inc.

Dawid, A. P. (2007a). Counterfactuals, hypotheticals and potential responses: A philosophical examination of statistical causality. In Causality and Probability in the Sciences, edited by F. Russo and J. Williamson. London: College Publications, Texts In Philosophy Series Vol. 5:503–532.

Dawid, A. P. (2007b). Fundamentals of statistical causality. (279).

Bibliography III

Dawid, A. P. and Didelez, V. (2010). Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview. *Statistics Surveys*, 4:184–231.

Guo, H. and Dawid, A. P. (2010). Sufficient covariates and linear propensity analysis. In Teh, Y. W. and Titterington, D. M., editors, *In Proceedings of the Thirteenth International Workshop on Artificial Intelligence and Statistics (AISTATS) 2010*, Journal of Machine Learning Research Workshop and Conference Proceedings 9, pages 281–288, Chia Laguna, Sardinia, Italy.

Lindley, D. V. (2002). Seeing and doing: the concept of causation. *International Statistical Review*, 70(2):191–197.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355.

Pearl, J. (1997). *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann.

Bibliography IV

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. Mathematical Modelling, 7:1393 – 1512.

Robins, J. (1987). Addendum to "a new approach to causal inference in mortality studies with sustained exposure periods-application to control of the healthy worker survivor effect". Computers & Mathematics with Applications, 14:923–945.

Robins, J. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. Biometrika, 79:321–324.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66:688–701.