Story for AIStats

1 Introduction

It is widely accepted that causal knowledge and statistical knowledge are distinct. At least two levels are common: statistics is concerned with *association* while causation is concerned with *consequences*; a distinction of this nature goes back at least to Hume [Morris and Brown, 2019], who is also noted for his argument that knowledge of the latter cannot be reduced to the former. Pearl [2009] has identified three level hierarchy of causal knowledge in contemporary work: first *association*, then *intervention* (analogous to Cartwright's *strategy*) and finally *counterfactuals*. Pearl argues that the types of things that can be known at higher levels subsumes what can be known at lower levels (e.g. all associational knowledge is a type of interventional knowledge), but lower levels do not subsume higher levels.

An apparently paradoxical feature of this three level hierarchy is that, though knowledge is claimed to flow only in one direction, we find that the first and third levels are both described by ordinary joint probability distributions. Counterfactual queries can be formulated as missing data problems, which are distinct from associational problems only due to the interpretations we assign to so-called *counterfactual random variables* or *potential outcomes*. Knowledge at the second level, on the other hand, is described by causal graphical models which are *not* joint probability distributions (in one treatment, they are introduced as indexed sets of joint probability distributions Pearl [2009]). Here is an apparently paradoxical feature of common approaches to causal inference: associational knowledge is distinguished from consequential knowledge in both interpretation and representation, while counterfactual knowledge – considered to subsume both – is distinguished from associational knowledge by interpretation only.

Statistical decision theory, introduced by Wald [1950], underpins much of modern machine learning. It introduced the fundamental notions of *loss* and *risk* to statistics and proided foundational theorems such as the *complete class theorem* which shows that all admissible decision rules are Bayesian decision rules for some prior. Key elements of statistical learning theory inherits heavily from statistical decision theory. While some descendants of statistical decision theory have grappled with the problem of causality [Lewis, 1981], SDT itself is regarded as a theory of statistical decision making and not of causality.

We show a surprising relationship between SDT and causal graphical models. We proceed in two steps: We note that a causal graphical model represents a relationship between probability measures on a given space and the consequences of a given set of actions. We then consider a modification of a standard statistical decision problem: suppose that, rather than being given a loss function that directly evaluates decisions, we are instead provided with a preference function over consequences of decisions that (following convention) we call a *utility*. The resulting problem is underspecified and provides no ordering over decisions. However, the type of relationship represented by a causal graphical model is then found to be precisely the type of object needed to fully specify the problem, and does so in a way that induces a regular statistical decision problem.

This motivates the definition of *causal statistical decision problems* (CSDPs). These relate to regular statistical decision problems (SDPs) in loose analogy with the way that model based reinforcement learning relates to model free reinforcement learning; while the former keeps track of both consequences and rewards/utililties of decisions, the latter "forgets about the consequences" and only works with reward/utiltiy.

> Is this true? There are substantial similarities between SDT and SLT, but I haven't found direct evidence of lineage in e.g. a citation from Valiant

CSDPs introduce the notion of *causal theories*. Causal theories represent relationships between probability measures and consequences and are a generalisation of causal graphical models. In Pearl's language, they represent the connection between associational knowledge and interventional knowledge; in Cartwright's, they connect associational knowledge with the consequences of strategies.

Thanks to the clarity of our approach, we are able to shed light on the questions raised in the second paragraph: we require a causal theory to bring knowledge from levels 1 to 2 of Pearl's hierarchy and we *also* require a causal theory to bring knowledge from level 3 to level 2. A joint distribution over counterfactuals can only answer interventional questions *given interventional assumptions* (we speculate that such assumptions may have been taken for granted). Associational knowledge is represented with probability distributions, knowledge of consequences with stochastic maps and relationships between the two with causal theories.

Choosing appropriate causal theories is a hard problem. Whether we build a causal theory with graphical models or Potential Outcomes (with additional assumptions), it is often the case that a nontrivial result rests on assumptions that are not obvious, generic or testable. Generic principles such as the bias-variance tradeoff have proved to be immensely powerful in the world of statistics, and we regard the question of whether there are generic principles that govern causal inference and what they may be to be one of the most important questions in the field.

We are primarily concerned with setting out a clear framework for reasoning about causal theories, and do not propose principles for constructing a causal theory in this paper. We are able to show a general negative result - causal theories that are symmetric over permutations of decisions cannot yield nontrivial decision rule orderings. We term this result "no causes in, no causes out" as it demonstrates that some causal knowledge is required at the outset if we hope for any nontrivial decision rules. Such asymmetric causal assumptions must be problem specific, so from the outset we cannot build causal theories on "problem neutral" assumptions alone.

> There's another half baked angle here, which is "what kinds of causal theories are represented by graphical models"? In particular, via the question of dominance we can consider causal theories to be related by three different types of randomisation. Also, if we examine marginal causal models, we note that they all represent causal theories that are related to a "nice" causal theory (in the sense that identification is straightforward) via two of these types of randomisation. It's half baked because I can't yet say a lot from there, save for the fact that the operation of randomisation seems more amenable to being generalised to a continuous version than DAGs do.

> I could also include the "free" results from statistical decision theory somewhere - complete class theorem, purification

2 Statistical Decision Problems and Causal Statistical Decision Problems

A statistical decision problem (SDP) poses the following scenario: suppose we have a set of "states of nature" Θ, a set of decisions D and a loss function $l : \Theta \times D \to \mathbb{R}$. For each state of nature $\theta \in \Theta$ there is an associated probability measure $\mu_\theta \in \Delta(\mathcal{E})$ where $(E, \mathcal{E})$ is some measurable space. Call the stochastic map $H : \theta \mapsto \mu_\theta$ a *statistical experiment*. Given a *decision strategy* $\pi : E \to \Delta(\mathcal{D})$, define the *risk* of π given state θ to be the expected loss of π in state θ. Specifically, $R : \Pi \times \Theta \to \mathbb{R}$ given by $R : (\pi, \theta) \mapsto \delta_\theta \curlyvee (H\pi \otimes \mathrm{Id}_\Theta)l$, where we make use of the product notation and copy map for brevity.

> Need a canonical measure on Θ; the coarsest measure rendering the evaluation maps measurable?

Supposing some unknown true state θ^*, we would ideally find a strategy π that minimises the risk in θ^*. Unfortunately, most statistical decision problems do not admit such strategies. Two alternative decision rules are available:

Given a measure $\xi \in \Delta(\Theta)$ called a prior, ξ-*Bayes decision rule* is a decision rule π_{Ba}^* such that the *Bayes risk* $R_\xi : \pi \mapsto \xi \curlyvee (H\pi \otimes \mathrm{Id}_\Theta)l$ is minimised:

$$\pi_{\mathrm{Ba}}^* \in \underset{\pi \in \Pi}{\arg\min}\, R_\xi(\pi) \tag{1}$$

A *minimax* decision rule π_{MM}^* minimises the worst-case risk. Unlike a Bayes rule, it does not invoke a prior:

$$\pi^*_{\text{Mm}} \in \arg\min_{\pi \in \Pi} \max_{\theta \in \Theta} R(\theta, \pi) \tag{2}$$

We emphasise here that we regard the set Θ as a "state of nature" or a "theory of nature" and not a "parameter set" - it is possible that for some $\theta \neq \theta'$ we have $\mu_\theta = \mu_{\theta'}$, a possibility not supported by the interpretation of Θ as a set of distribution parameters. If there were a decision strategy that minimised the loss in every state, such a strategy would clearly minimise the loss in the true state.

Our representation of statistical experiment is slightly different to, for example, Le Cam [1996], who introduces statistical experiments as an ordered collection of probability measures. Both representations do the same job, and the representation as a map makes for a clearer connection with causal statistical decision problems.

Formally, we define an SDP as the tuple $\langle \Theta, E, D, H, l \rangle$ where Θ, E and D are measurable sets, H is a stochastic map $\Theta \to \Delta(\mathcal{E})$ and l a measurable function $E \to \mathbb{R}$. We leave implicit the set Π of decision strategies $E \to \Delta(\mathcal{D})$ and $\mathbb{R}$, the codomain of l.

This is a very bare bones exposition of the theory of SDPs, and for more details we refer readers to Ferguson [1967].

Observe that a statistical decision problem supplies a loss l that tells us immediately how desirable a pair $(\theta, d) \in \Theta \times D$ is. In many areas it is more typical to talk about how desirable the *consequences* of a decision are than how desirable a (state, decision) pair is. If the set of possible consequences of a decision is denoted by a set F, let the desirability of an element $f \in F$ be given by a utility function $u : F \to \mathbb{R}$; "utility" being a very conventional term for such a desirability function. Given such a u, the tuple $\langle \Theta, E, D, H, u \rangle$ is an ill-posed problem; we want to evaluate the desirability of decisions D (or decision strategies π), but we have no means of connecting decisions with consequences F. An obvious move is to introduce for each state of nature θ a *consequence map* $\kappa_\theta : D \to \Delta(\mathcal{F})$; let C be the map $\theta \mapsto \kappa_\theta$. We can then define the *causal risk* $S : \Pi \times \Theta \to \mathbb{R}$ by $R : (\pi, \theta) \mapsto \delta_\theta \curlyvee (H\pi \otimes ID\Theta)Cu$, and Bayes and minimax risks are defined by obvious analogy.

For each state $\theta \in \Theta$, the Markov kernel

$$T_\theta := \quad \begin{array}{c} \boxed{H} \!\!-\! E \\ \langle \delta_\theta \mathbin{\!-\!} \\ \boxed{C} \!\!-\! F \end{array} \tag{3}$$

<div style="display: flex;">
<div style="flex: 3;">

Is sufficient to compute the causal risk. Thus we can, in general, replace Θ, H and C with the *causal theory* $\mathcal{T} : \{T_\theta | \theta \in \Theta\}$. A causal statistical decision problem (CSDP) is therefore a tuple $\langle \mathcal{T}, E, F, D, u \rangle$. We can recover $H' : \mathcal{T} \to \Delta(\mathcal{E})$ by $H' := (\text{Id}_E \otimes *_F)$ and $C' : \mathcal{T} \to (D \to \Delta(\mathcal{F}))$ by $C' : (*_E \otimes \text{Id}_F)$ such that $\delta_\theta H = T_\theta H'$ and $\delta_\theta C = T_\theta C'$. Noting that a statistical decision problem can be defined for an arbitrary set Θ, given a CSDP and recovering letting $l := Cu$ we can canonically induce an SDP $\langle \mathcal{T}, E, D, H, l \rangle$

</div>
<div style="flex: 1; background: orange;">

This is highly nonobvious, depends on the work of Jacobs, and is only known to be true for finite sets E and F. On the other hand, there's a very intuitive graphical proof.

</div>
</div>

Call the set $\mathcal{T} := \{(H_\theta, C_\theta) | \theta \in \Theta\}$ the *causal theory*. Note that

	SDPs	CSDPs
State of the world	Θ	$\mathcal{T}$, causal theory
Observation space	E	E
Result space	-	F
Decisions	D	D
Given preferences	$\ell : \Theta \times D \to \mathbb{R}$	$u : F \to \mathbb{R}$
Loss in a given state	$\ell(\theta, \cdot), \theta \in \Theta$	$\kappa u(\cdot), (\kappa, \mu) \in \mathcal{T}$

Table 1: Comparison of SDPs and CSDPs

We develop causal statistical decision problems (CSDPs) inspired by statistical decision problems (SDPs) of Wald [1950]. CSDPs differ from SDPs in that our preferences (i.e. utility or loss) are known less directly in former case. We show that every SDP can be represented by a CSDP and that the converse is sometimes but not always possible. We show that an analogue of the fundamental *complete class theorem* of SDPs applies to the class of CSDPs that can be represented by SDPs, but whether such a theorem applies more generally is an open question.

3

Following [Ferguson, 1967], we consider SDPs and CSDPs to represent normal form two person games. At the most abstract level the games represent the options and possible payoffs available to the decision maker, and this representation allows us to compare the two types of problem. In their more detailed versions, CSDPs and SDPs differ in their representation of the state of the world and in the type of function that represents preferences. These differences are summarised in Table 1.

Definition 2.1 (Normal form two person game). A normal form game is a triple $\langle \mathcal{S}, A, L \rangle$ where $\mathcal{S}$ and A are arbitrary sets and $L : \mathcal{S} \times A \to [0, \infty)$ is a loss function.

The set $\mathcal{S}$ is a set of possible states that the environment may occupy and A is a set of actions the decision maker may take. The decision maker seeks an action in A that minimises the loss L. Generally there is no action that minimises the loss for all environment states. A minimax solution is an action that minimises the worst case loss: $a_{mm}^* = \arg\min_{a \in A}[\sup_{s \in \mathcal{S}} L(s, a)]$.

If the set $\mathcal{S}$ is equipped with a σ-algebra $\mathcal{S}$ and a probability measure $\xi \in \Delta(\mathcal{S})$ which we will call a "prior", a Bayes solution minimizes the expected risk with respect to ξ: $a_{ba}^* = \arg\min_{a \in A} \int_{\mathcal{S}} L(s, a)\xi(ds)$.

Definition 2.2 (Admissible Action). Given a normal form two person game $\langle \mathcal{S}, A, L \rangle$, an action $a \in A$ is *strictly better* than $a' \in A$ iff $L(s, a) \leq L(s, a')$ for all $s \in \mathcal{S}$ and $L(s_0, a) < L(s_0, a')$ for some $s_0 \in \mathcal{S}$. If only the first holds, then a is as good as a'. An *admissible action* is an action $a \in A$ such that there is no action strictly better than A.

Definition 2.3 (Complete Class). A class C of decisions is a *complete class* if for every $a \notin C$ there is some $a' \in C$ that is strictly better than a.

C is an *essentially complete* class if for every $a \notin C$ there is some $a' \in C$ that is as good as a.

A statistical decision problem represents a normal form two-person game where the available actions are *decision functions* that output a decision given data, the states of the environment are associated with probability measures on some measurable space and we assume a loss expressing preferences over decisions and states is known.

Definition 2.4 (Statistical Experiment). A *statistical experiment* relative to a set Θ, a measurable space $(E, \mathcal{E})$ and a map $m : \Theta \to \Delta(\mathcal{E})$ is a multiset $\mathcal{H} = \{\mu_\theta | \theta \in \Theta\}$ where $\mu_\theta := m(\theta)$. The set Θ indexes the "state of nature".

Definition 2.5 (Statistical Decision Problem). A statistical decision problem (SDP) is a tuple $\langle \Theta, (\mathcal{H}, m), D, \ell \rangle$. $\mathcal{H} \subset \Delta(\mathcal{E})$ is a statistical experiment relative to states Θ, space $(E, \mathcal{E})$ and map $m : \Theta \to \Delta(\mathcal{E})$, D is the set of available decisions with some σ-algebra $\mathcal{D}$ and $\ell : \Theta \times D \to \mathbb{R}$ is a loss function where $\ell(\theta, \cdot)$ is measurable with respect to $\mathcal{D}$ and $\mathcal{B}(\mathbb{R})$.

Denote by $\mathcal{J}$ the set of stochastic decision functions $E \to \Delta(\mathcal{D})$. For $J \in \mathcal{J}$ and $\mu_\theta \in \mathcal{H}$, the risk $R : \Theta \times \mathcal{J} \to [0, \infty)$ is defined as $R(J, \theta) = \int_D \ell(\theta, y)\mu_\theta J(dy)$. The triple $\langle \Theta, \mathcal{J}, R \rangle$ forms a two player normal form game.

The loss function ℓ expresses preferences over general (state, decision) pairs. It may be the case that our preferences are most directly known over future states of the world - we know which results of our decisions are desirable and which are undesirable, which we represent with a *utility function*. In this case, if we are to induce preferences over the possible decisions, that we have a model that is more informative than a statistical experiment. In particular, we require each state of nature to be associated with both a distribution over the given information and a map from decisions to distributions over results - we call this map a *consequence*, and the object that pairs a distribution and a consequence with each state of the world a *causal theory*.

Definition 2.6 (Consequences). Given a measurable result space $(F, \mathcal{F})$ and a measurable decision space $(D, \mathcal{D})$, a Markov kernel $\kappa : D \to \Delta(\mathcal{F})$ is a *consequence mapping*, or just a *consequence*.

Definition 2.7 (Causal state). Given a consequence $\kappa : D \to \Delta(\mathcal{F})$, a measurable observation space $(E, \mathcal{E})$ and some distribution $\mu \in \Delta(\mathcal{E})$, the pair (κ, μ) is a *causal state* on E, D and F. We refer to κ as the consequence and μ as the observed distribution.

In many cases the observation space E and the results space F might coincide. However, these spaces are defined by different aspects of the given information: the former is fixed by what observations are available and the latter by which parts of the world are relevant to the investigator's preferences (see Theorems **??** and **??**), and there is not a clear reason to insist that these spaces should always be the same.

Definition 2.8 (Causal Theory). A causal theory $\mathfrak{T}$ is a set of causal states sharing the same decision, observation and outcome spaces. We abuse notation to assign the "type signature" $\mathfrak{T} : E \times D \twoheadrightarrow F$ for a causal theory with observed distributions in $\Delta(\mathcal{E})$ and consequences of type $D \to \Delta(\mathcal{F})$. The causal states of a theory $\mathfrak{T}$ may be associated with a master set of states Θ, but in contrast to a statistical experiment this is not necessary to define the basic associated decision problem.

Definition 2.9 (Causal Statistical Decision Problem). A causal statistical decision problem (CSDP) is a triple $\langle \mathfrak{T}, D, u \rangle$. $\mathfrak{T}$ is a causal theory on $D \times E \twoheadrightarrow F$, D is the decision set with σ-algebra $\mathcal{D}$ and $u : F \to \mathbb{R}$ is a measurable utility function expressing preference over the results of decisions.

Define the canonical loss $L : \mathfrak{T} \times D \to \mathbb{R}$ by $L : (\kappa, \mu), y \mapsto -\mathbb{E}_{\gamma\kappa}[u]$. This change conforms with the conventions that utilities are maximised while losses are minimised.

Given a decision function $J \in \mathcal{J}$ and $(\kappa, \mu) \in \mathfrak{T}$, we define the risk $R : \mathfrak{T} \times \mathcal{J} \to [0, \infty)$ by $R(\kappa, \mu, J) := L((\kappa, \mu), \mu J)$. The triple $\langle \mathfrak{T}, \mathcal{J}, R \rangle$ is a normal form two person game.

The loss and the utility differ in that the loss expresses per-state preferences while the utility expresses state independent preferences. While we choose the loss to be a particular function of the utility here, it is possible to allow losses to be a more general class of functions of the utility and state without altering the preference ordering of a CSDP under minimax or Bayes decision rules. Given arbitrary $f : \mathfrak{T} \to \mathbb{R}$, define $l : \mathfrak{T} \times D \to \mathbb{R}$ by $l : (\kappa, \mu, y) \mapsto af(\kappa, \mu) + b\mathbb{E}_{\delta_y\kappa}[u]$. We can define a loss (relative to f) $L : \mathfrak{T} \times \Delta(\mathcal{D}) \to [0, \infty]$ by

$$L((\kappa, \mu), \gamma) := \mathbb{E}_\gamma[l(\kappa, \mu, \cdot)] \tag{4}$$
$$= af(\kappa, \mu) - b\mathbb{E}_{\gamma\kappa}[u] \tag{5}$$
$$\tag{6}$$

For $(\kappa, \mu) \in \mathfrak{T}$, $\gamma \in \Delta(\mathcal{D})$ and $a \in \mathbb{R}$, $b \in \mathbb{R}^+$.

A common example of a loss of the type above is the *regret*, which takes $a = b = 1$ and $f(\kappa, \mu) = \sup_{\gamma' \in \Delta(\mathcal{D})} \mathbb{E}_{\gamma'\kappa}[u]$. Because expected utility preserves preference orderings under positive affine transformations, the ordering of preferences given a particular state is not affected by the choices of a, b and f, nor is the Bayes ordering of preferences given some prior ξ over $\mathfrak{T}$. While it may be possible to formulate decision rules for which the choices of a, b and f do matter, we will take these properties as sufficient to allow us to choose $a = 0$ and $b = 1$. More general classes of loss are of interest. *Regret theory*, for example, is a straightforward generalisation of the losses discussed here and is a prominent alternative to expected utility theory [**?**].

There are obvious similarities between SDPs and CSDPs: both have the same high level representation as a two person game which is arrived at by taking the expectation of a loss with respect to a decision function. In fact, if we consider two decision problems to be the same if they have the same representation as a two player game, we find that CSDPs are a special case of SDPs.

Theorem 2.10 (CSDPs are a special case of SDPs). *Given any CSDP $\alpha = \langle \mathfrak{T}, D, u \rangle$ with two player game representation $\langle \mathfrak{T}, \mathcal{J}, R \rangle$, there exists an SDP $\langle \mathfrak{T}, (\mathcal{H}, m), D, \ell \rangle$ with the same representation as a two player game.*

Proof. Let $m : \mathfrak{T} \to \mathcal{H}$ be defined such that $m : (\kappa, \mu) \mapsto \mu$ for $(\kappa, \mu) \in \mathcal{H}$. Define $\ell : \mathfrak{T} \times D \to \mathbb{R}$ by $\ell : ((\kappa, \mu), y) \mapsto -\mathbb{E}_{\delta_y\kappa}[u]$. Let $R'((\kappa, \mu), J) = \mathbb{E}_{\mu J}[\ell(\theta, \cdot)]$. Then

$$R'((\kappa, \mu), J) = -\int_D \mathbb{E}_{\delta_y\kappa}[u]\mu J(dy) \tag{7}$$

$$= -\int_D \int_F u(x)\kappa(y; dx)\mu J(dy) \tag{8}$$

$$= -\int_F u(x)\mu J\kappa(dx) \tag{9}$$

$$= R((\kappa, \mu), J) \tag{10}$$

$$\square$$

The converse is not true, as the set Θ in an SDP is of an arbitrary type and may not be a causal theory. However, it is possible for any SDP with environmental states Θ to find a CSDP with causal theory $\mathfrak{T}$

such that the games represented by each decision problem are related by a surjective map $f : \Theta \to \mathcal{T}$ which associates each state of nature with a causal state. We call such a map a *reduction* from an SDP to a CSDP.

Definition 2.11 (Reduction)**.** Given normal form two person games $\alpha = \langle \mathcal{S}^\alpha, A, L^\alpha \rangle$ and $\beta = \langle \mathcal{S}^\beta, A, L^\beta \rangle$, $f : \mathcal{S}^\alpha \to \mathcal{S}^\beta$ is a *reduction* from α to β if, defining the image $f(\mathcal{S}^\alpha) = \{f(\theta) | \theta \in \mathcal{S}^\alpha\}$, we have $\langle \mathcal{S}^\beta, A, L^\beta \rangle = \langle f(\mathcal{S}^\alpha), A, L^\alpha \circ (f \otimes I_A) \rangle$.

Theorem 2.12 (SDP can be reduced to a CSDP)**.** *Given any SDP $\langle \Theta, (\mathcal{H}, m), D, \ell \rangle$ represented as the game $\alpha = \langle \Theta, \mathcal{J}, R \rangle$, there exists a CSDP $\langle \mathcal{T}, D, u \rangle$ represented as the game $\beta = \langle \mathcal{T}, \mathcal{J}, R' \rangle$ such that there is some reduction $f : \Theta \to \mathcal{T}$ from α to β.*

Proof. Take $\mathcal{H} \subset \Delta(\mathcal{E})$ and define $f : \Theta \to \Delta(\mathcal{E}) \times \Delta(\mathcal{B}(\mathbb{R}))^D$ by $f : \theta \mapsto (y \mapsto \delta_{l(\theta,y)}, \mu_\theta)$. Noting that $y \mapsto \delta_{l(\theta,y)}$ is a Markov kernel $D \to \Delta(\mathcal{B}(\mathbb{R}))$, the image $f(\Theta)$ is a causal theory $E \times D \twoheadrightarrow \mathbb{R}$. Consider the CSDP $\langle f(\Theta), D, -I_{(\mathbb{R})} \rangle$. Then, letting R' denote the risk associated with this theory

$$R'((\kappa, \mu), J) = -\int_{\mathbb{R}} \int_D (-x) \delta_{l(\theta,y)}(dx) \mu_\theta J(dy) \tag{11}$$

$$= \int_D l(\theta, y) \mu_\theta J(dy) \tag{12}$$

$$= R(\Theta, J) \tag{13}$$

$\square$

The fundamental *complete class theorem* of SDPs establishes that there are no decision rules that dominate the set of all Bayes rules under some regularity assumptions. By theorem 2.10, this must also be true of CSDPs.

Theorem 2.13 (Complete class theorem (CSDP))**.** *Given any CSDP $\alpha := \langle \mathcal{T}, D, u \rangle$ with two player game representation $\langle \mathcal{T}, \mathcal{J}, R \rangle$, if $|\mathcal{T}| < \infty$ and $\inf_{J \in \mathcal{J}, (\kappa, \mu) \in \mathcal{H}} R((\kappa, \mu), J) > -\infty$, then the set of all Bayes decision functions is a complete class for α and the set of all admissible Bayes decision functions is a minimal complete class for α.*

Proof. By theorem 2.10, there exists an SDP β such that α and β have the same representation as a two player game. By assumption, β has a finite set of states and a risk function that is bounded below. Therefore the Bayes rules on α are a complete class and admissible Bayes rules are a minimal complete class for the problem $\langle \mathcal{T}, \mathcal{J}, R \rangle$ [Ferguson, 1967]. $\square$

References

Thomas S. Ferguson. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, July 1967. ISBN 978-1-4832-2123-6.

L. Le Cam. Comparison of Experiments - A Short Review.pdf. *IMS Lecture Notes - Monograph Series*, 30, 1996. URL `https://www.fastmailusercontent.com/mail-attachment/f2132276u33762/%3CB868E7FAD8F0C04FA0207113AD7C77A9%40ausprd01.prod.outlook.com%3E/Comparison%20of%20Experiments%20-%20A%20Short%20Review.pdf?u=9a2e41ab&a=195efd739077a3e70251567c111372cb67a6dcf0`.

David Lewis. Causal decision theory. *Australasian Journal of Philosophy*, 59(1):5–30, March 1981. ISSN 0004-8402. doi: 10.1080/00048408112340011. URL `https://doi.org/10.1080/00048408112340011`.

William Edward Morris and Charlotte R. Brown. David Hume. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition, 2019. URL `https://plato.stanford.edu/archives/sum2019/entries/hume/`.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2 edition, 2009.

Abraham Wald. *Statistical decision functions*. Statistical decision functions. Wiley, Oxford, England, 1950.