A Short Note on the Equivalence of the Ontic and the Epistemic View on Data Imprecision for the Case of Stochastic Dominance for Interval-Valued Data

Georg Schollmeyer GEORG.SCHOLLMEYER@STAT.UNI-MUENCHEN.DE
Department of Statistics, LMU Munich

Abstract

In the context of the analysis of interval-valued or set-valued data it is often emphasized that one has to carefully distinguish between an epistemic and an ontic understanding of set-valued data. However, there are cases, for which an ontic and an epistemic view do still lead to exactly the same results of the corresponding data analysis. The present paper is a short note on this fact in the context of the analysis of stochastic dominance for interval-valued data.

Keywords: Relational Data Analysis, Stochastic Dominance, Partially Ordered Set, Interval Order, Ontic and Epistemic View, Cautious Data Completion

1. Introduction

There are at least two different types of views on interpreting the data imprecision that is possibly inherent in interval-valued or set-valued data, (c.f., e.g., [8] for an exemplification of this disambiguation):

- In the epistemic view, a set-valued data point represents an imprecise observation of a precise, but not directly observable data point of interest. One has partial knowledge about this precise, but not observed data point. One only knows that the unknown precise data point is a member of a set-valued observed data point.

- Opposed to this, in the ontic view, a set-valued data point is understood as a precise observation of something that is 'imprecise' only in the sense that we do not observe $\mathbb{R}^p$-valued data, but set-valued data. The observed set is set-valued by nature and there are no distinguished elements in the observed set and there is actually no real imprecision at all. As an example, think of the lifetimes of persons. Say, Joseph Haydn lived from 1732 to 1809. At every time point from 1732 to 1809 Haydn was alive, but there is no special distinguished time point within this period. For such a prolonged data 'point' one still has a natural notion of order: One can say that Haydn definitely 'lived before' Robert Schuhmann (1810-1856). On the other hand, Haydn did not definitely live 'before' or 'after' Wolfgang Amadeus Mozart (1756-1791), he wos born before Mozart and survived his dead. So it is natural to speak here of an ontic type of a partial order. Beyond the partial order inherent in such type of interval data, there is of course further partial numeric information. Within this paper that is concerned with stochastic dominance, we will only rely on the aspects of the underlying partial order.

For both views on set-valued data, different approaches for handling these types of data where proposed in the literature and it is often emphasized, especially within the community of imprecise probabilities, that it is important to not confuse these concepts in the first place. Furthermore, such aspects of data imprecision often propagate to subtle issues of model imprecision: Imprecisely observed data often make statistical models only partially identified and thus the true underlying parameter cannot be estimated consistently and there is only a set (usually called identified set or identification region) of possible parameters that could be identified via the distribution of all involved random variables that can actually be observed. In this situation of partial identification[1], the question arises, if the non-identified true parameter or the whole identification region is the object of interest, which roughly corresponds to an epistemic versus an ontic understanding. Examples of the epistemic understanding can be found e.g., in [16] while the ontic approach is used e.g., in [15, 6]. Interestingly, in the literature on partial identification which treats such situations of partially identified models, there seems to be no explicit reference to an epistemic or an ontic view, however, a discussion of the disambiguation that does not use the words epistemic or ontic can be found for example in [28].

However one might judge about the usefulness of an explicit disambiguation between an ontic and an epistemic view, it can sometimes be intriguing to apply methods that are designed in an ontic fashion to situations of epistemic data- or model imprecision. One example is the so-called Set-loss Region (cf., [25]), an identification region that was developed in the context of partially identified linear regression models under interval-valued

1. For an introduction to partial identification, see, e.g., [29].

response variables. This region treats the interval-valued responses in an ontic fashion. But as the analysis shows, this misuse of the ontic view leads to an identification region with acceptable statistical properties and with a clear relation to the epistemic view. For details, see [25] (especially paragraph 3.2 and 4.4, as well as appendix A), where also the so-called marrow region and the so-called collection region, both with an epistemic underpinning, are discussed.[2]

Also in many applications of machine learning to interval-valued data, the epistemic view could be well-motivated, but it is imaginable that an ontic approach could be still more effective in a certain statistical sense and at the same time the ontic approach may still be conceptually relatable to the epistemic view. As an example, think of the widely used 'ontic' approach of the Hausdorff-distance for measuring the distance between set-valued data, of course often without any attempt of scrutinization of the used understanding of data imprecision, compare e.g., the discussion in [30]. The present paper is specifically concerned with differences between the ontic and the epistemic view w.r.t. data imprecision (not w.r.t. e.g. model imprecision[3]). One specific aspect of the epistemic view in this situation is the fact that the epistemic line of reasoning is often in the spirit of the so-called cautious data completion (cf., [3, paragraph 7.8, p.181]):

"First look at every possible precise data point that is compatible with the observed interval-valued data and then apply a "classical" method to the set of all such possible data-completions." But this seemingly straightforward and innocently looking reasoning may possibly have the following three Achilles' heels:

1. By applying a classical precise method in the second step, it is not appreciated that the method has to deal with potential data points and not with the actual true data point. Thus, the method cannot see that this or that specific potential data point should possibly be taken not as seriously as if one would know for sure that the given data point is in fact the actually observed true data point.

2. In dividing the procedure into two steps one handles things as if one would divide a logical analysis into two steps, but conceptually, classical statistics could not be more far away from the picture of deductive reasoning, statistics in its classical form is all about only 'controlling'[4] the statistical behavior of a method, it was never about and will never be about 'controlling' truth or argument, in fact there is no such thing as an inductive logic of inference:

   "And the success of science is not based upon rules of induction, but depends upon luck, ingenuity, and the purely deductive rules of critical argument.... Induction, i.e. inference based on many observations, is a myth. It is neither a psychological fact, nor a fact of ordinary life, nor one of scientific procedure." Popper [23, S.53].

3. From a purely mathematical point of view, in dividing a method into two steps that know not from each other, one is vulnerable to any kind of ineffectiveness and also ill-posedness issues (which are in fact present e.g., in the context of linear regression under partial identification, c.f., the discussion in [25, paragraphap 5 and appendix A] and [21]) :

   "When solving a problem of interest, do not solve a more general problem as an intermediate step. Try to get the answer that you really need but not a more general one." Vapnik [31, S.477]

Of course, on the other hand, the 'misuse' of an ontic procedure, concatenated with an epistemic reinterpretation, is also some kind of a two-step procedure and thus also in danger of suffering from the third point. For example, there exists a big amount of literature (cf., e.g., [4, 5]) utilizing random set theory as an ontic approach for constructing confidence regions for the identified set of a partially identified statistical model. In typical cases of partial identification, the fact that the model is only partially identified comes from the fact that certain variables cannot be observed. But this means that the object of interest is still the unknown true parameter and not the identified set, at least if one believes in the 'true parameter' at all. (For a motivation of an interest in the whole identification region, see [28].) Thus, one would actually be more interested in a confidence region for the true, unknown parameter, which seems to be difficult to obtain directly within a random set approach. (Of course, a confidence region for the identified set is also a conservative confidence region for the true parameter, but it is in fact typically conservative, cf., e.g. [4, p. 778] or [16, Lemma 1].)

2. The further splitting into the marrow region and the collection region is not due to a further disambiguation within the epistemic view, it is only due to a conceptual disambiguation of the underlying model understanding as a structural or as a descriptive model (cf., also [21]), which is not of relevance in this paper.
3. But note that often aspects of model imprecision are induced by data imprecision. As already mentioned, for example imprecisely observed data often lead to partially identified statistical models.
4. With 'controlling', I simply mean here the control of the error probabilities e.g. in a classical hypothesis test. Of course, it is only a controlling w.r.t. the theoretical term probability, if one takes Cournot's principle (cf., [7, p.78]) and Poppers fallibilism (cf., [22, paragraph 68]) seriously, then one is far away from really controlling things in empirical terms.

However one is prepared to face the above considerations, the present note is concerned with a situation, where the disambiguation between an epistemic and an ontic view does not make a difference with respect to the results one obtains by applying the different views on data imprecision.

2. First Order Stochastic Dominance

In this paper, we deal with the notion of first order stochastic dominance under interval-valued data. The concept of first order stochastic dominance plays an important role in a huge variety of disciplines like for example in decision theory (cf., e.g., [19]), welfare economics (cf., e.g., [1, 2]), portfolio analysis (cf., e.g., [17]), nonparametric item response theory (cf., e.g., [24]), medicine (cf., e.g., [18]), toxicology (cf., e.g., [9]) or psychology (cf., e.g., [20]).

One typical simple example is the analysis of income poverty for example w.r.t. two different subpopulations. Think for example of the distribution of the income in two different countries, which can be formalized with two random variables X and Y. The idea behind first order stochastic dominance is to say that X is stochastically (weakly) smaller than Y if the probability of X taking high values (i.e. $P(X \geq c)$) is always smaller than or equal to the probability of Y taking high values (i.e. $P(Y \geq c)$) independent of the chosen threshold c, usually called 'poverty-line'). If, in the sense of Sens capability approach (cf., [27]), one wants to jointly analyze more 'dimensions' of poverty, the analysis is more difficult. One aspect here is that dimensions like e.g. education do not have a cardinal scale of measurement, in fact, one can argue that the dimension education is only of a partially ordered scale of measurement. (It appears natural to say that one person has a lower education than another person only if she followed the same or a comparable educational path, but stopped earlier.) Additionally, it is difficult to compare different dimensions with different scales of measurement. A natural way to do a stochastic poverty analysis is then a relational analysis: Define person x as less poor than person y (i.e. $x \geq y$) if she is less poor w.r.t. every dimension of poverty. The concept of stochastic dominance then translates to the analysis of so-called upsets instead of events of the form $X \geq c$:

Definition 1 *Let $(V, \leq)$ be a partially ordered set that is additionally equipped with a σ-algebra $\mathscr{A}' \subseteq 2^V$. A set $A \subseteq V$ is called an upset (w.r.t. $(V, \leq)$), if we have*

$$\forall x, y \in V : x \in A \;\&\; x \leq y \implies y \in A.$$

Let now $(\Omega, \mathscr{A}, P)$ be a probability space and let $X, Y : \Omega \longrightarrow V$ be two V-valued random variables. We say that X is (weakly) stochastically smaller than Y (i.e. $X \leq_{SD} Y$) if for every $\mathscr{A}'$-measurable upset $A \subseteq V$ it holds that

$$P_X(A) \leq P_Y(A).$$

Remark 2 *The above definition can be intuitively explained by saying:*

- *Every (measurable) upset A can be understood as a reasonable concretization of the term non-poor: Formally declare all values $x \in A$ as non-poor and all values $x \notin A$ as poor. Then, such a set A is a reasonable concretization of the term non-poor, if it holds that for every non-poor value $x \in A$ and every value y that is better than x or equal to x w.r.t. every dimension (i.e.: $x \leq y$) also y is declared as non-poor. This is exactly the property of being an upset w.r.t. $\leq$ and beyond this, if we have only the relation $\leq$, there is no further constraint for a reasonable concretization of the term non-poor.*
- *Then, X is (weakly) stochastically smaller than Y iff for all reasonable concretizations of the term non-poor, the probability for X being declared as non-poor is smaller than or equal to the probability of Y being declared as non-poor. Or short: X is (weakly) stochastically smaller than Y iff the probability of being non-poor is smaller for X compared to Y, independently of the (reasonable) concretization of the term non-poor.*

Remark 3 *To give a little more intuition of the concept of stochastic dominance, an equivalent characterization is given by saying that X is (weakly) stochastically smaller than Y iff for every bounded, isotone and measurable function u (think of an utility function) we have*

$$\mathbb{E}(u \circ X) \leq \mathbb{E}(u \circ Y),$$

which could be interpreted as: 'Whatever the utility function u, the expected utility for X is always smaller than or equal to the expected utility for Y'.

Remark 4 *If $(V, \leq)$ is linearly ordered, then the upsets of V are simply the upper-closed intervals of the form $[c, \infty] := \{x \in V \mid x \geq c\}$ and $]c, \infty] := \{x \in V \mid x > c\}$, respectively. Note further that for checking stochastic dominance in empirically observed samples, one needs only to look at observed values c and not at all arbitrary $c \in \mathbb{R}$. Compared to the linearly ordered case, in the partially ordered case the set of all upsets of a partially ordered set typically is very large and explicitly checking every upset is illusory. In the worst case, there are essentially 2^n upsets for empirically observed samples of total size n. However, there are efficient characterizations in terms of linear programs for checking stochastic dominance for empirically observed samples, as well as a first approach to do statistical inference in this situation of poset-valued random*

variables, see, e.g., [26]. In the following, we are explicitly only concerned with detecting stochastic dominance for empirically observed samples, meaning that P_X and P_Y are replaced by empirical analogues. Some aspects of statistical inference are only briefly touched in Section 5.

3. Analyzing First Order Stochastic Dominance Under Epistemic Data-imprecision

Consider now the following situation: Assume, for simplicity, we are interested in the simple analysis of income poverty, but now with the additional difficulty that the income is not precisely observed. Instead, for every respondent we have only an interval $[l,u]$ and we only know that the true income lies in the interval $[l,u]$. This situation is in fact a realistic scenario, in surveys like e.g. [13] one often asks for the income firstly with a direct request, but for the non-responders, one adds a follow up question about the income in a categorized question-design to decrease the non-response rate. Then, according to the above disambiguation of the two understandings of data imprecision, for every view, a natural way to proceed would suggest itself. (Of course, conceptually, the epistemic data view is adequate here, because there is a precise, but unknown income.)

Epistemic remedy: In the epistemic view, it appears natural to compute stochastic dominance for every potential data point that is compatible with the observed intervals and to say that $X \leq_{SD} Y$ if stochastic dominance is valid for every potential data point.

Ontic remedy: In the ontic view one can firstly define a relation for interval-valued data points. The most straightforward definition would be the application of an interval order (cf., e.g., [12]) in the sense that for two intervals $[l,u]$ and $[l',u']$ one defines $[l,u] \leq [l',u'] :\iff u \leq l'$.

Of course, also other orderings for intervals are possible, cf., e.g., [10, p.1366], but the interval order from above is the 'most conservative' one. If one has not a relational kind of analysis in mind, then another way to proceed would be for example a metric approach. One could for example introduce an ontic notion of distance between intervals, e.g., the Hausdorff-distance. However, with respect to first order stochastic dominance, the notion of order is the essential aspect and not the notion of distance. Because of this, we proceed here with a relational ontic analysis and the interval order from above. The following theorem now exactly expresses the fact, that essentially both views on data imprecision will lead to the same results of the analysis. The involved mapping Φ_A exactly describes here the process of the data-completion.

Theorem 5 *Let $\leq$ be the linear order of the reals and let $(I,\leq_I)$ be an interval order, meaning that there exist functions $l,u : I \longrightarrow \mathbb{R}$ that represent $(I,\leq_I)$ in the sense of*

$$\forall x,y \in I : x \leq_I y \iff u(x) \leq l(y).$$

For $A \subseteq I$ define the homomorphism

$$\Phi_A : I \longrightarrow \mathbb{R} : x \mapsto \begin{cases} u(x) & \text{if } x \in A \\ l(x) & \text{else} \end{cases}.$$

Then we have

1. *If $A \subseteq I$ is an upset w.r.t. $(I,\leq_I)$, then $\Phi_A(A) := \{\Phi_A(x) \mid x \in A\}$ is an upset w.r.t. $(\Phi_A(I), \leq \cap \Phi_A(I) \times \Phi_A(I))$.*

2. *Let furthermore*

$$\Phi : I \longrightarrow \mathbb{R} : x \mapsto \Phi(x) \in [l(x),u(x)]$$

be an arbitrary data completion. If $\Phi(A)$ is an upset w.r.t. $(\Phi(I), \leq \cap \Phi(I) \times \Phi(I))$, then $\{x \in I \mid \exists z \in A : \Phi(z) = \Phi(x)\} = \Phi^{-1}(\Phi(A))$ is an upset in $(I,\leq_I)$.

Proof

1. Let A be an upset in $(I,\leq_I)$, let $a \in \Phi_A(A)$ and let $b \in \Phi_A(I)$ with $b \geq a$. We have to show that $b \in \Phi_A(A)$: We have $b = \Phi_A(y) \geq a = \Phi_A(x)$ for appropriate $x \in A$ and $y \in I$. It follows $y \in A$, because if $y \notin A$ then necessarily $y \not\geq_I x$ and we would have $b = \Phi_A(y) = l(y) < u(x) = \Phi_A(x) = a$ which contradicts the assumption $b \geq a$. Thus, $y \in A$ and therefore $b = \Phi_A(y) \in \Phi_A(A)$.

2. Let $\Phi : I \longrightarrow \mathbb{R} : x \mapsto \Phi(x) \in [l(x),u(x)]$ be an arbitrary data completion and let $\Phi(A)$ be an upset in $(\Phi(I), \leq \cap \Phi(I) \times \Phi(I))$, let $x \in \Phi^{-1}(\Phi(A))$ or equivalently $\Phi(x) \in \Phi(A)$ and let $y \geq_I x$. We have to show that $y \in \Phi^{-1}(\Phi(A))$ or equivalently that $\Phi(y) \in \Phi(A)$: Since Φ is a consistent w.r.t. l and r, we have $\Phi(y) \geq l(y) \geq u(x) \geq \Phi(x)$. Since $\Phi(x) \in \Phi_(A)$ and since $\Phi_(A)$ is an upset, we have $\Phi(y) \in \Phi(A)$.

■

4. A Brief Exemplification of the Theorem

The first part of the theorem states that for an ontic type analysis, every upset A in the interval order $(I,\leq_I)$ associated to the interval-valued data gives rise to a specific data completion Φ_A for which $\Phi_A(A)$ is an upset in $\leq \cap \Phi_A(I) \times \Phi_A(I)$. In particular this means that if we have

$X \not\leq_{SD} Y$ in an ontic analysis, there exists an upset $A \subseteq I$ with

$$P_X(A) > P_Y(A),$$

or equivalently[5]

$$P_X(\Phi_A^{-1}(\Phi_A(X))) > P_Y(\Phi_A^{-1}(\Phi_A(X))),$$

or

$$P_{\Phi_A \circ X}(\Phi_A(A)) > P_{\Phi_A \circ Y}(\Phi_A(A)).$$

The first part of the theorem then implies that $\Phi_A(A)$ is an upset in $\leq \cap \Phi_A(I) \times \Phi_A(I)$, which means, that also an epistemic analysis would establish $X \not\leq_{SD} Y$.

The second part of the theorem exactly expresses the reversed implication: If we have $X \not\leq_{SD} Y$ in an epistemic type analysis, then there exists a data completion Φ and an upset $\Phi(A)$ in $\leq \cap \Phi_A(I) \times \Phi_A(I)$ with

$$P_{\Phi \circ X}(\Phi(A)) > P_{\Phi \circ Y}(\Phi(A)),$$

which is equivalent to the statement

$$P_X(\Phi^{-1}(\Phi(A))) > P_Y(\Phi^{-1}(\Phi(A))),$$

and since the second part of the theorem states that $\Phi^{-1}(\Phi(A))$ is an upset in $(I, \leq_I)$, it follows that also an ontic type analysis would yield $X \not\leq_{SD} Y$.

5. General Conclusion

In this paper we have established the fact that for first order stochastic dominance under interval-valued data, the epistemic and the ontic view on data imprecision still lead to the same result. This result is of course more or less obvious, especially in the case of a linearly ordered set, but it establishes the possibility to look at the problem both from an epistemic and an ontic view which allows to utilize techniques from one view to facilitate the analysis w.r.t. the other view. Let us briefly elaborate a little bit on this by starting with an obvious observation: For an epistemic data analysis of first order stochastic dominance and empirically observed samples for analyzing e.g. income-poverty between two countries, it is easy to get the most extreme data completions by assigning the lower bounds of the observed intervals to one country and the upper bounds to the other country. Thus, if one has to deal with this simple situation, one could easily apply classical methods for univariate first order stochastic dominance to these extreme data completions. In more complicated situations of e.g. multidimensional poverty-analysis one has to rely on methods for detecting stochastic dominance like that developed in [26]. How one incorporates possible data imprecision of certain dimensions, i.e., if one replaces the intervals with the extreme values or if one simply uses the corresponding interval-order does not play any role w.r.t. the result.

But beyond this obviousness, the ontic remedy additionally allows for imposing further modeling assumptions: If the coarsening process leading to the interval-valued data is not coarsening at random, but if one can assume at least that the coarsening process is the same for X and Y, then one can conclude that the distribution of the true but unknown data within the observed intervals is the same for X and Y. This legitimates to define observed intervals $[l,u]$ and $[l',u']$ with the same endpoints as equivalent (i.e. $[l,u] \leq_I [l',u']$ and $[l',u'] \leq_I [l,u]$). Note that this assumption does not mean that the coarsening process of X or Y does not depend one the values of X or Y, it only means that the dependence is the same for X and Y. With this assumption which is clearly weaker than coarsening at random one would get an analysis, that is generally more decisive than an epistemic analysis, which is unable to incorporate this modelling assumption. (Of course, technically, in an epistemic analysis one can constrain the possible data completions to that ones, for which the precise potential data that correspond to identical observed intervals are identical, too. This would lead to the same result as the ontic procedure, but is conceptually not so much in the spirit of the epistemic view, because there is no assumption that the precise data values are identical for identical observed intervals, one only assumes that the distribution of the precise data within the observed intervals is identical, which is a statement that is more in the fashion of the ontic view.)

Another point that makes the 'ontic misuse' attractive is the aspect of inference. Generally, in the multidimensional case, statistical inference is very difficult already for the case of precisely observed data. In this situation, useful empirical analogues that characterize stochastic dominance like e.g.,

$$\sup_{A \text{ upset}} \hat{P}_X(A) - \hat{P}_Y(A)$$

are not distribution free like in the continuous univariate case. Statistical inference could then be done with permutation tests. Within the epistemic view, the cautious data

5. Note that for an upset A we have $\Phi_A^{-1}(\Phi(A)) = A$ or equivalently $\Phi_A(x) \in \Phi_A(A) \iff x \in A$. The second direction of this equivalence is obvious and for the first direction assume that $\Phi_A(x) \in \Phi_A(A)$. Then there exists $y \in A$ with $\Phi_A(x) = \Phi_A(y)$. If $x \notin A$, because of $y \in A$ we would have $I(x) = \Phi_A(x) = \Phi_A(y) = u(y)$ and thus $y \leq_I x$, and thus $x \in A$ because A was an upset in $(I, \leq_I)$. But this contradicts the assumption $x \notin A$ and thus the assumption $x \notin A$ was false and we have $x \in A$.

completion would then require to do a permutation test for every possible data completion, which would be computationally intractable. But due to the theorem, it is actually not needed to look at the cautious data completion, it is enough to do a permutation test on the ontic level. Furthermore, a rough statistical analysis of the statistical complexity of the problem in terms of the Vapnik-Chervonenkis dimension[6] of the underlying family of all upsets of $(V, \leq)$ could be easily applied on the ontic level. It turns out that the Vapnik-Chervonenkis dimension is exactly the width[7] of the underlying poset $(V, \leq)$ of the ontic approach, see [26, Section 5.2.1]. Because of the isomorphism between $(V, \leq)$ and

$$\bigcap_{\Phi \text{ data completion}} \leq_\Phi$$

where $x \leq_\Phi y \iff \Phi(x) \leq \Phi(y)$, also within an epistemic analysis the highest Vapnik-Chervonenkis dimension as Φ ranges over every possible data completion is thus also identical to the width of $(V, \leq)$. This means that one can obtain large deviation bounds for the test statistic of interest (independently of the view on data imprecision). Since the width of $(V, \leq)$ could be arbitrarily high, the test statistic can easily become badly behaved. In this situation, one can try to regularize the problem like indicated in [26, Section 5.3.1] (The approach for regularization described therein treats the case of precisely observed data, but it can be adopted to the case of imprecisely observed data). So one can say that statistical analysis and regularization within an epistemic understanding are very easy only due to the somehow trivial theorem from above.

From a more general view, in contrast to the specific case of first order stochastic dominance where the extreme data completions are easy to obtain, the ontic remedy is often computationally easier to handle than the epistemic one. One prominent example is the computation of the cautious data completion for the variance under interval-valued data. Ferson et al. [11] showed that computing the upper bound for the variance under interval-valued data is NP hard. Compared to this, in an ontic fashion it is straightforward to define a measure of dispersion for interval-valued data by replacing distances in $\mathbb{R}^p$ by e.g., the Hausdorff-distance between intervals or sets. Of course, one would become one single number for the dispersion that seems to be not translatable into an epistemic view.

Concerning the epistemic view, note that ironically, the variance as a measure of dispersion is so widely used because of its nice algebraic/analytical properties, which makes the variance easy to handle. This is of special interest in the light of the fact that the variance as a 'measure of dispersion' is not the only reasonable choice. One very underrated alternative measure of dispersion with good structural properties[8] is Gini's mean difference, cf., [14]. This measure can be related to linear moments and thus seems to be computationally more accessible w.r.t. the cautious data completion than an L^2-type measure like the variance: While the computation of the cautious data completion for Gini's mean difference can be done by solving linear programming problem (which has polynomial smoothed complexity), for computing the variance, the most straightforward approach would be a quadratic programming procedure. But for the upper bound of the variance one would have to solve a quadratic program with a quadratic form that is not negative-definite, which indicates the already mentioned fact that computing the upper bound is in fact NP hard. It cannot be emphasized enough that in the first place, a typical statistical data analysis is often mostly silent about which exact 'concept' of e.g. dispersion has to be used, and if this silence is adequate, then there is much room for using alternatives that are computationally realizable.

To conclude, one can say that in spite of the important warning to not confuse the epistemic and the ontic view, one should always keep one's eyes open, if, within the competition of ideas, the ontic remedy can say something about the epistemic or vice versa, notwithstanding the fact that we actually need more direct methods and analyses instead of 'two-step crutches'.

Acknowledgments

The author would like to thank the three anonymous reviewers for their very helpful comments and the LMU Mentoring program, supporting young researchers for providing financial support.

References

[1] Channing Arndt, Roberta Distante, M. Azhar Hussain, Lars Peter Østerdal, Pham Lan Huong, and Maimuna Ibraimo. Ordinal welfare comparisons with multiple discrete indicators: A first order dominance approach and application to child poverty. *World Development*, 40(11):2290–2301, 2012.

6. Informally speaking, the Vapnik-Chervnenkis of a family of sets is a measure of the statistical complexity of the family of sets in terms of the large deviation behaviour of a supremum type statistic over this family of sets. For more details, see, e.g., [32].
7. The width of a poset is the maximal cardinality of a subset of pairwise incomparable elements.
8. One astonishing property of Gini's mean difference is that for two samples or distributions X and Y with the same median and for which the upper tail of X (i.e., the distribution above the median) is stochastically smaller (w.r.t. first order stochastic dominance) than that of Y, and the lower tail of X is stochastically larger than that of Y, Gini's mean difference for X is always smaller than Gini's mean difference for Y, as one would intuitively expect. Opposed to this, the variance generally violates this homomorphism-property.

[2] Channing Arndt, Nikolaj Siersbæk, and Lars Peter Østerdal. Multidimensional first-order dominance comparisons of population wellbeing. WIDER Working Paper 2015/122, 2015. URL http://hdl.handle.net/10419/129471. accessed 15.05.2019.

[3] Thomas Augustin, Gero Walter, and Frank P.A. Coolen. Statistical inference. In Thomas Augustin, Frank P A Coolen, Gert de Cooman, and Matthias C M Troffaes, editors, *Introduction to Imprecise Probabilities*. Wiley, 2014.

[4] Arie Beresteanu and Francesca Molinari. Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814, 2008.

[5] Arie Beresteanu, Ilya Molchanov, and Francesca Molinari. Sharp identification regions in models with convex moment predictions. *Econometrica*, 79(6): 1785–1821, 2011.

[6] Victor Chernozhukov, Han Hong, and Elie Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284, 2007.

[7] Antoine Augustin Cournot. *Exposition de la théorie des chances et des probabilités*. L. Hachette, 1843.

[8] Ines Couso and Didier Dubois. Statistical reasoning with set-valued information: Ontic vs. epistemic views. *International Journal of Approximate Reasoning*, 55(7):1502–1518, 2014. Special issue: Harnessing the information contained in low-quality data sources.

[9] Ori Davidov and Shyamal Peddada. Testing for the multivariate stochastic order among ordered experimental groups with application to dose–response studies. *Biometrics*, 69(4):982–990, 2013.

[10] Thierry Denoeux. Extending stochastic ordering to belief functions on the real line. *Information Sciences*, 179(9):1362–1376, 2009.

[11] Scott Ferson, Lev Ginzburg, Vladik Kreinovich, Luc Longpré, and Monica Aviles. Computing variance for interval data is NP-hard. *SIGACT News*, 33(2): 108–118, June 2002.

[12] Peter C Fishburn. Intransitive indifference with unequal indifference intervals. *Journal of Mathematical Psychology*, 7(1):144–149, 1970.

[13] GESIS - Leibniz - Institut fur Sozialwissenschaften. *ALLBUS compact (2015): Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*. 2015. GESIS Datenarchiv, Köln. ZA5241 Datenfile Version 1.1.0.

[14] Corrado Gini. Variabilità e mutabilità. *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi*, 1912.

[15] Joel L. Horowitz and Charles F. Manski. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*, 95(449):77–84, 2000.

[16] Guido W. Imbens and Charles F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004.

[17] Timo Kuosmanen. Efficient diversification according to stochastic dominance criteria. *Management Science*, 50(10):1390–1406, 2004.

[18] Moshe Leshno and Haim Levy. Stochastic dominance and medical decision making. *Health Care Management Science*, 7(3):207–215, 2004.

[19] Haim Levy. *Stochastic Dominance: Investment Decision Making under Uncertainty*. Springer, 2015.

[20] Haim Levy and Moshe Levy. Experimental test of the prospect theory value function: A stochastic dominance approach. *Organizational Behavior and Human Decision Processes*, 89(2):1058–1081, 2002. doi:10.1016/S0749-5978(02)00011-0.

[21] Maria Ponomareva and Elie Tamer. Misspecification in moment inequality models: back to moment equalities? *The Econometrics Journal*, 14(2):186–203, 2011.

[22] Karl Popper. *The logic of scientific discovery*. Hutchinson, 1959.

[23] Karl Popper. *Conjectures and refutations: The growth of scientific knowledge*. Routledge, 2014.

[24] Hartmann Scheiblechner. A unified nonparametric IRT model for d-dimensional psychological test data (d-ISOP). *Psychometrika*, 72(1):43, 2007.

[25] Georg Schollmeyer and Thomas Augustin. Statistical modeling under partial identification: Distinguishing three types of identification regions in regression analysis with interval data. *International Journal of Approximate Reasoning*, 56:224–248, 2015.

[26] Georg Schollmeyer, Christoph Jansen, and Thomas Augustin. Detecting stochastic dominace for poset-valued random variables as an example of linear programming on closure systems. Technical Report 209, Department of Statistics, LMU Munich, 2017. URL http://www.statistik.uni-muenchen.de/forschung/technical_reports/index.html.

[27] Amartya Sen. *Development as Freedom*. Oxford University Press, 1999.

[28] Jörg Stoye. Statistical inference for interval identified parameters. In Thomas Augustin, Frank Coolen, Matthias Troffaes, and Serafin Moral, editors, *ISIPTA'09: Proc. Sixth Int. Symp. on Imprecise Probability: Theories and Applications*, pages 395–404. Citeseer, 2009.

[29] Elie Tamer. Partial identification in econometrics. *Annual Review of Economics*, 2(1):167–195, 2010.

[30] Lev V Utkin and Anatoly I Chekh. A new robust model of one-class classification by interval-valued training data using the triangular kernel. *Neural Networks*, 69:99–110, 2015.

[31] Vladimir N. Vapnik. *Estimation of dependences based on empirical data. Empirical inference science: Afterword of 2006 / Vladimir Vapnik*. Springer, 2006. ISBN 0387308652.

[32] Vladimir N. Vapnik. *Estimation of Dependences Based on Empirical Data*. Springer, 2006.