Measures in Databases and Data Mining

Bassem Sayrafi[1], Dirk Van Gucht[1*], and Marc Gyssens[2]

[1] Computer Science Department,Indiana University
Bloomington, IN 47405-4101, USA
{bsayrafi,vgucht}@cs.indiana.edu
[2] Department WNI, University of Limburg (LUC)
Universitaire Campus, B-3590 Diepenbeek, Belgium
marc.gyssens@luc.ac.be

Abstract. We present a framework to study the properties of measures as they occur in various areas of databases and data mining such as aggregation in queries, measurement of data uniformity, and frequency calculation. The framework is a generalization of the theory of mathematical measures. In particular, our framework is built on principles that relax the additivity principle for mathematical measures. Besides using our framework to classify measures, we derive general bounds and rules they must satisfy. By considering the analogue of first and second derivatives of functions, in our case the first and second finite differences of measures, we obtain inference systems that allow us to reason about constraints that exist between data objects relative to measurements.

* The first two authors were supported by NSF Grant IIS-0082407.

1 Introduction

Measures frequently occur in database and data mining applications. In database queries and in data cube computations, measures such as `count`, `max`, `min`, `sum`, and `average` are commonly used [18]. In query processing and optimization, histograms, which are synthesized frequency objects naturally emerge [25]. In data mining applications such as the frequent item sets problem, the frequency of occurrences set of items in a collection of baskets is central [1].[3] Other areas of information processing where measures are used include information retrieval [6], web searching [8], and bibliometrics [13].

The main contribution of our paper is the introduction of an axiomatic framework to define and study the properties of such additive measures. Conceptually, our framework is to additive measures what first-order logic is to queries. To the database and data mining theory community, our approach offers a platform for the systematic study of a significant class of measures. To the database and data mining community at large, our approach can be interpreted for specific measures, and provides a tool to study their properties.

Our framework for measures is inspired by the mathematical framework for measures over finite sets [10]. In that framework, given a finite set S, a nonnegative function $\mathbf{m} : 2^S \rightarrow \mathbf{R}$, is called a measure if for each pair Y and Z of *disjoint* subsets of S (i.e., $Y \cap Z = \emptyset$), we have that

$$\mathbf{m}(Y \cup Z) = \mathbf{m}(Y) + \mathbf{m}(Z).$$

This property is called additivity. It can easily be seen that, as a consequence, mathematical measures also satisfy the following two properties:

$$\mathbf{m}(X) \leq \mathbf{m}(X \cup Y); \tag{1}$$

$$\mathbf{m}(Y \cup Z) + \mathbf{m}(Y \cap Z) = \mathbf{m}(Y) + \mathbf{m}(Z). \tag{2}$$

The first property is called *isotonicity*, the second *modularity*.[4]

[3] Most of these measures are of an additive nature: the measure of a data object can be estimated by summing the measures of its sub-objects.

[4] The modularity property clearly is an instance of the generalized inclusion-exclusion principle [20].

Whereas mathematical measures satisfy the pair of inequality (1) and equality (2), many measures only satisfy weaker conditions. Consider, for example, the $\texttt{max}$ function defined over finite sets of nonnegative numbers. This function clearly has the isotonicity property since $\texttt{max}(X) \leq \texttt{max}(X \cup Y)$, but it does not have the modularity property, since, in general, it is not the case that $\texttt{max}(Y \cup Z) + \texttt{max}(Y \cap Z) = \texttt{max}(Y) + \texttt{max}(Z)$ (take $Y = \{1,2\}$ and $Z = \{1,3\}$ to see this). Thus $\texttt{max}$ is *not* a mathematical measure. However, as can be easily shown, $\texttt{max}$ satisfies the weaker condition

$$\texttt{max}(Y \cup Z) + \texttt{max}(Y \cap Z) \leq \texttt{max}(Y) + \texttt{max}(Z).$$

which we will call *submodularity.*

Given the insights above, a more general framework for measures could be based on the isotonicity and submodularity properties.[5] Such a framework would be too narrow, however. Too see this, consider the $\texttt{min}$ function over finite sets. Obviously, a framework for measures that encompasses $\texttt{max}$ should also encompass $\texttt{min}$. Unfortunately, $\texttt{min}$ is neither isotone, nor submodular. However, $\texttt{min}$ is *anti-isotone*, since $\texttt{min}(X) \geq \texttt{min}(X \cup Y)$, and $\texttt{min}$ is "anti-" submodular (henceforth called *supermodular*), i.e.,

$$\texttt{min}(Y \cup Z) + \texttt{min}(Y \cap Z) \geq \texttt{min}(Y) + \texttt{min}(Z).$$

Therefore, functions that are anti-isotone and supermodular will also be considered measures.

In summary, at the core of our framework for measures are the concepts of isotonicity, anti-isotonicity, submodularity, and supermodularity for nonnegative (finite) set-based functions. These concepts and their properties are summarized in Section 2. It turns out that many measures occurring in databases and data mining can be accommodated in our framework. This will be pursued in Section 3 and will serve as justification for the appropriateness of the approach we chose.

Many other useful properties of measures can be derived in the framework. As a simple example, for an isotone measure $\mathcal{M}$, it is

[5] Observe that it follows from the previous discussion that each mathematical measure (on finite sets) will be a measure in our framework, but clearly not the other way around.

the case that $\max(\mathcal{M}(Y), \mathcal{M}(Z)) \leq \mathcal{M}(Y \cup Z) \leq \mathcal{M}(Y) + \mathcal{M}(Z)$. This property and others are derived in Section 4. Arguably, the most interesting properties derived there provide upper and lower bounds on the measures of sets in terms of the measures of certain of their subsets. The main proposition in the section relates measures to average functions, which can be shown to be increasing and concave for isotone measures, and decreasing and convex for anti-isotone measures.

A natural issue that arises when considering measures is their *rate of change*. For a measure $\mathcal{M}$, we consider the rate of change at a set X, with displacement Y, to be the quantity $\mathcal{M}(X \cup Y) - \mathcal{M}(X)$. Observe that there is a direct analogy to the theory of *finite differences* [17]. Our rate of change corresponds to the first finite difference. A high value for this difference signals strong independence (anti-correlation) between X and $X \cup Y$ (relative to $\mathcal{M}$), whereas a low value indicates strong interdependency (correlation) between X and $X \cup Y$.) For instance, when the rate is zero, i.e., $\mathcal{M}(X) = \mathcal{M}(X \cup Y)$, one has that the measure at $X \cup Y$ is entirely determined by the measure at X, i.e, X completely determines Y relative to $\mathcal{M}$. In addition to considering the rate of change of $\mathcal{M}$, it turns out that studying the rate of change of the rate of change of $\mathcal{M}$ (i.e., the second finite difference) is also significant. The first and the second rate of change functions of measures, which we call *measure differentials*, are studied in Sections 5. In particular we derive a series of inequalities that hold between measure differentials.

These inequalities become useful when we consider them in the context of the study of *constraints* on measures. The most natural constraints occur when certain measure differentials are zero. We consider this issue in Section 6. In that section, using the inequalities on measure differentials, we obtain inference rules for such constraints. The inference systems thus obtained are very similar in form to inference systems that have appeared in the literature, such as the Armstrong axioms for functional dependencies in relational databases [4]. We conclude Section 6 by interpreting our constraints for various measures discussed in Section 3. In Section 7, finally, we exploit these interpretations to prove the completeness of some of our inference systems for measure constraints.

2 Measures

In this section, we define various kinds of measures and relate them to each other. We also make a comparison with probabilistic measures [24].

2.1 Basic definitions and properties

In the rest of the paper, S denotes a finite set, $\mathcal{M}$ and $\mathcal{N}$ denote nonnegative real functions over 2^S, and the variables U, V, X, Y, and Z range over 2^S. Furthermore, XY is short for $X \cup Y$.

Definition 1. The function $\mathcal{M}$ is

isotone	if $\mathcal{M}(X) \leq \mathcal{M}(XY)$;
anti-isotone	if $\mathcal{M}(X) \geq \mathcal{M}(XY)$;
submodular	if $\mathcal{M}(YZ) + \mathcal{M}(Y \cap Z) \leq \mathcal{M}(Y) + \mathcal{M}(Z)$;
supermodular	if $\mathcal{M}(YZ) + \mathcal{M}(Y \cap Z) \geq \mathcal{M}(Y) + \mathcal{M}(Z)$.

If $\mathcal{M}$ is isotone or anti-isotone, then $\mathcal{M}$ is called a *weak measure*. If $\mathcal{M}$ is isotone and submodular, respectively anti-isotone and supermodular, then $\mathcal{M}$ is called a *measure*.

The following, obvious proposition justifies our terminology.

Proposition 1. *A measure is also a weak measure.*

The following proposition (the proof of which is straightforward) will allow us to primarily focus on isotone weak measures and measures only.

Proposition 2. *Let $\mathcal{M}$ and $\mathcal{N}$ be related as follows as follows:*

$$\mathcal{N}(X) = \mathcal{M}(S) + \mathcal{M}(\emptyset) - \mathcal{M}(X).$$

Then

1. *$\mathcal{N}$ is isotone if and only if $\mathcal{M}$ is anti-isotone, and vice-versa;*
2. *$\mathcal{N}$ is submodular if and only if $\mathcal{M}$ is supermodular, and vice-versa.*

Submodular and supermodular functions [15] have been studied extensively because they are the discrete analogues of concave and convex functions, respectively, and because they emerge in many combinatorial optimization problems.

The following proposition gives a useful characterization of measures.

Proposition 3. *1. The function $\mathcal{M}$ is an isotone measure if and only if $\mathcal{M}(XYZ) + \mathcal{M}(X) \leq \mathcal{M}(XY) + \mathcal{M}(XZ)$.*
2. The function $\mathcal{M}$ is an anti-isotone measure if and only if $\mathcal{M}(XYZ) + \mathcal{M}(X) \leq \mathcal{M}(XY) + \mathcal{M}(XZ)$.

Proof. We only prove the first statement. By applying Proposition 3, the second statement easily follows.

We first consider the *if*. Thus, suppose the condition stated above holds. The isotonicity condition can be obtained from it by equating Z with Y. Similarly, the submodularity condition can be obtained by equating X with $Y \cap Z$.

We now turn to the *only if*. Thus, suppose $\mathcal{M}$ is both isotone and submodular. Because of submodularity,

$$\mathcal{M}(XYZ) + \mathcal{M}(XY \cap XZ) \leq \mathcal{M}(XY) + \mathcal{M}(XZ).$$

Because of isotonicity, $X \subseteq XY \cap XZ$ implies $\mathcal{M}(X) \leq \mathcal{M}(XY \cap XZ)$. Therefore,

$$\mathcal{M}(XYZ) + \mathcal{M}(X) \leq \mathcal{M}(XY) + \mathcal{M}(XZ).$$

Submodularity *alone* does not imply isotonocity, and hence not the condition in Proposition 3. Arguably, the best known examples of functions that are submodular and not isotone are *cut functions* of graphs [15]. Let $G = (S, E)$ be a finite directed graph. For $V \subseteq S$, define the *cut* of V, denoted $\delta(V)$, as follows:

$$\delta(V) = \{(v, w) \in E \mid v \in V \text{ and } w \notin V\}.$$

The *cut function* $f : 2^S \to \mathbf{R}^{\geq 0}$ of the graph G is defined as follows[6]:

$$f(V) = |\delta(V)|.$$

[6] Usually the cut function of a graph is defined in the context of a *capacity function*. This function associates with each edge in the graph a weight (its capacity). Here we use the capacity function that associates with each edge the value 1.

It can be shown that f is a submodular function, i.e.,

$$f(VW) + f(V \cap W) \leq f(V) + f(W).$$

However, f is not necessarily isotone. For example, let $S = \{a, b\}$ and $E = \{(a, b)\}$. Then $f(\{a\}) = 1$, but $f(\{a, b\}) = 0$. Consequently, this function f is not a measure in the sense of Definition 1.

2.2 Relationship to probabilistic measures

A finite *probability* space is defined as a triple $(S, 2^S, p)$, where S is the sample space, 2^S the set of events, and $p : 2^S \to [0, 1]$ a function such that

1. $p(S) = 1$; and
2. $p(YZ) = p(Y) + p(Z)$ if $Y \cap Z = \emptyset$.

In particular, it follows that p is a probabilistic measure. It is an easy exercise that probabilistic measures are always isotone and submodular, and therefore isotone measures according to Definition 1. Thus, each probabilistic measure is a submodular measure.

The opposite is of course not true. However, for a non-constant submodular measure $\mathcal{M}$, we can define the function

$$p_{\mathcal{M}}(X) = \frac{\mathcal{M}(X) - \mathcal{M}(\emptyset)}{\mathcal{M}(S) - \mathcal{M}(\emptyset)}.$$

Clearly, $p_{\mathcal{M}} : 2^S \to [0, 1]$ is a monotone function satisfying

1. $p_{\mathcal{M}}(S) = 1$; and
2. $p_{\mathcal{M}}(YZ) \leq p_{\mathcal{M}}(Y) + p_{\mathcal{M}}(Z)$ if $Y \cap Z = \emptyset$.

Thus, with each submodular function we can associate a function that has certain characteristics of probabilistic measures.

3 Measures in database and data mining applications

In the following subsections, we describe a variety of application areas in databases and data mining where (weak) measures occur naturally. We identify these and fit them in the framework specified in Section 2.

In the area of databases, we consider aggregation functions, and, in the area of data uniformity in relational databases, we consider the probability-based Gini index and the Shannon entropy measure. In the area of data mining, we focus on measures occurring in the context of the frequent item set problem.

3.1 Databases - aggregation functions

Computations requiring aggregation functions occur frequently in database applications such as query processing, datacubes [18], and spreadsheets. Among the most often used aggregation functions are `count`, `sum`, `min`, `max`, and *order statistics*. Each of these functions operates on finite sets. Except in the case of `count`, we assume that the sets on which they work consist of nonnegative integers. We show that all of these functions are either weak measures or measures.

1. Let $\mathtt{count}(X)$ be the cardinality of X. By Proposition 3, `count` is an isotone measure, since

$$\mathtt{count}(XYZ) + \mathtt{count}(X) \leq \mathtt{count}(XY) + \mathtt{count}(XZ).$$

2. Let $\mathtt{sum}(X) = \sum_{x \in X} x$, for $X \neq \emptyset$, and $\mathtt{sum}(\emptyset) = 0$. By Proposition 3, `sum` is an isotone measure, since

$$\mathtt{sum}(XYZ) + \mathtt{sum}(X) \leq \mathtt{sum}(X \cup Y) + \mathtt{sum}(XZ).$$

3. Let $\mathtt{max}(X)$ be the largest integer in X, for $X \neq \emptyset$, and let $\mathtt{max}(\emptyset)$ be the smallest element in S. Now, either $\mathtt{max}(XYZ) = \mathtt{max}(XY)$ or $\mathtt{max}(XYZ) = \mathtt{max}(XZ)$. In the former case, it follows from $\mathtt{max}(X) \leq \mathtt{max}(XZ)$ that

$$\mathtt{max}(XYZ) + \mathtt{max}(X) \leq \mathtt{max}(XY) + \mathtt{max}(XZ).$$

   In the latter case, the inequality above follows from $\mathtt{max}(X) \leq \mathtt{max}(XY)$ and $\mathtt{max}(XYZ) = \mathtt{max}(XZ)$. By Proposition 3, `max` is an isotone measure.
4. Let $\mathtt{min}(X)$ be the smallest integer in X, for $X \neq \emptyset$, and let $\mathtt{min}(\emptyset)$ be the largest element in S. With an argument running along the same lines as in the previous case, one can easily see that `min` is an anti-isotone measure.

5. Order statistics are used to determine the ith smallest element of a set. For example, the second order statistic, denoted $\texttt{min2}(X)$, returns the second smallest element in X. More precisely, $\texttt{min2}(X) = \texttt{min}(X - \{\texttt{min}(X)\})$. Clearly, min2 is anti-isotone. However, it is not supermodular: the supermodularity condition in Definition 1 fails for $Y = \{1, 10, 11\}$ and $Z = \{2, 10, 11\}$. In general, order statistics are weak measures.

Aggregate functions derived from the above are not necessarily measures or weak measures however. For example, consider the average function defined by $\texttt{avg}(X) = \texttt{sum}(X)/\texttt{count}(X)$, if X is nonempty, and $\texttt{avg}(\emptyset) = 0$. This function is not isotone ($X = \{1\}$ and $Y = \{0, 1\}$ in the defining condition in Definition 1) and not anti-isotone ($X = \{1\}$ and $Y = \{1, 2\}$). From this example, it follows, that, in general, the quotient of two measures is not necessarily a measure or even a weak measure. Similarly, the variance and median functions are not weak measures.

3.2 Databases—data uniformity

Consider the values occurring in an attribute of a relation. These values can occur uniformly (e.g., the values "male" and "female" in the gender attribute of a census), or skewed (e.g., the values for the profession attribute in the same census). Measuring these degree of uniformity can influence how data of these types are stored or processed, or both. When data are numeric, a common way to measure uniformity is the variance statistic. This statistic computes the average of the distances between data values and their average. To measure data uniformity for categorical data, however, it is more useful to consider probability-based measures. Here, we consider two such measures: the *Gini index* and the *Shannon entropy measure*. The Gini index was introduced in economics to study the distribution of incomes [16].[7] The Shannon entropy measure was introduced to measure uniformity in communication data [27]. Unlike variance, these functions are specified in terms of probability distributions defined over the data sets.

[7] A closely related measure is the Simpson diversity measure was introduced to study the concept of diversity in ecologies and is defined as 1 minus the Gini index [29].

The Gini index Let ϱ be a nonempty finite relation over the relation schema S, and let p be a probability distribution over ϱ satisfying $p(t) \neq 0$ for all t in ϱ. For $X \subseteq S$, define p_X to be the marginal probability distribution of p on X. Thus, for $x \in \pi_X(\varrho)$, $p_X(x) = \sum_{\{t\in\varrho|t[X]=x\}} p(t)$. Now define

$$\mathcal{G}(X) = \sum_{x\in\pi_X(\varrho)} p_X(x)(1 - p_X(x)) = 1 - \sum_{x\in\pi_X(\varrho)} p_X^2(x).$$

Notice that $\mathcal{G}(\emptyset) = 0$. The function $\mathcal{G}$ is known as the *Gini index*.

We first show that $\mathcal{G}$ is isotone. Therefore, let X and Y be subsets of S, and put $Z = XY$. Since $X \subseteq Z$, we have that, for each $x \in X$, $p_X(x) = \sum_{z\in\pi_Z(\varrho_x)} p_Z(z)$, where $\varrho_x = \sigma_{X=x}(\varrho)$. Therefore,

$$p_X^2(x) = \Big[\sum_{z\in\pi_Z(\varrho_x)} p_Z(z)\Big]^2 \geq \sum_{z\in\pi_Z(\varrho_x)} p_Z^2(z).$$

Hence,

$$\begin{aligned}\mathcal{G}(X) = 1 - \sum_{x\in\pi_X(\varrho)} p_X^2(x) &\leq 1 - \sum_{x\in\pi_X(\varrho)} \sum_{z\in\pi_Z(\varrho_x)} p_Z^2(z)\\ &= 1 - \sum_{z\in\pi_Z(\varrho)} p_Z^2(z) = \mathcal{G}(Z) = \mathcal{G}(XY).\end{aligned}$$

Observe that the double summation collapses to a single summation because $\{\pi_Z(\varrho_x) \mid x \in \pi_X(\varrho)\}$ is a partition of $\pi_Z(\varrho)$.

We next show that $\mathcal{G}$ is submodular. Let $U = Y \cap Z$ and $V = YZ$. We have to prove that $\mathcal{G}(U) + \mathcal{G}(V) \leq \mathcal{G}(Y) + \mathcal{G}(Z)$. Thereto, let $u \in \pi_U(\varrho)$. The contribution of u to both terms of the left-hand side of the inequality is

$$\begin{aligned}&p_U(u)(1 - p_U(u)) + \sum_{v\in\pi_V(\varrho_u)} p_V(v)(1 - p_V(v))\\ &\quad= p_U(u) - p_U^2(u) + \sum_{v\in\pi_V(\varrho_u)} (p_V(v) - p_V^2(v_u))\\ &\quad= p_U(u) - \Big[\sum_{v\in\pi_V(\varrho_u)} p_V(v)\Big]^2 + \sum_{v\in\pi_V(\varrho_u)} p_V(v) - \sum_{v\in\pi_V(\varrho_u)} p_V^2(v)\\ &\quad= 2p_U(u) - \Big[\Big[\sum_{v\in\pi_V(\varrho_u)} p_V(v)\Big]^2 + \sum_{v\in\pi_V(\varrho_u)} p_V^2(v)\Big].\end{aligned}$$

The contribution of u to both terms of the right-hand side of the inequality is

$$\begin{aligned}
&\sum_{x\in\pi_X(\varrho_u)} p_X(x)(1-p_X(x)) + \sum_{y\in\pi_Y(\varrho_u)} p_Y(y)(1-p_Y(y)) \\
&\quad = \sum_{x\in\pi_X(\varrho_u)} \left[p_X(x) - p_X^2(x)\right] + \sum_{y\in\pi_Y(\varrho_u)} \left[p_Y(y) - p_Y^2(y)\right] \\
&\quad = \sum_{x\in\pi_X(\varrho_u)} p_X(x) - \sum_{x\in\pi_X(\varrho_u)} p_X^2(x) + \sum_{y\in\pi_Y(\varrho_u)} p_Y(y) - \sum_{y\in\pi_Y(\varrho_u)} p_Y^2(y) \\
&\quad = 2p_U(u) - \left[\sum_{x\in\pi_X(\varrho_u)} p_X^2(x) + \sum_{y\in\pi_Y(\varrho_u)} p_Y^2(y)\right] \\
&\quad = 2p_U(u) - \left[\sum_{x\in\pi_X(\varrho_u)} \left[\sum_{v\in\pi_V(\varrho_x)} p_V(v)\right]^2 + \sum_{y\in\pi_Y(\varrho_u)} \left[\sum_{v\in\pi_V(\varrho_y)} p_V(v)\right]^2\right].
\end{aligned}$$

Since each term in the quadratic expansion of

$$\sum_{x\in\pi_X(\varrho_u)} \left[\sum_{v\in\pi_V(\varrho_x)} p_V(v)\right]^2 + \sum_{y\in\pi_Y(\varrho_u)} \left[\sum\nolimits_{v\in\pi_V(\varrho_y)} p_V(v)\right]^2$$

occurs, in an injective way, in the quadratic expansion of

$$\left[\sum_{v\in\pi_V(\varrho_u)} p_V(v)\right]^2 + \sum_{v\in\pi_V(\varrho_u)} p_V^2(v),$$

the submodularity inequality (Definition 1) holds relative to u. Since u was arbitrary, the submodularity inequality holds.

Hence, the Gini index is an isotone measure in our setting.

The Shannon entropy measure Define

$$\mathcal{H}(X) = -\sum_{x\in\pi_X(\varrho)} p_X(x)\log p_X(x).$$

The function $\mathcal{H}$ is known as the *Shannon entropy measure*. It was proved in [21, 14] that $\mathcal{H}$ is isotone and submodular. A simple proof can be found in [12]. Hence, the Shannon entropy measure is an isotone measure in our setting.

3.3 Data mining—frequent item sets

An important problem in data mining is discovering frequent item sets. In this problem, a set of baskets is given. Each basket contains a set of items. In practice, the items may be products sold at a grocery store, and baskets correspond to items bought by its customers. The frequent item sets problem is to find the item sets that occur frequently together within the baskets. In the example, these frequent item sets can guide how items are placed strategically in the store. The notion of frequency associated with this problem turns out to be an anti-isotone measure in our setting. The anti-isotonicity principle known as the apriori rule is at the heart of the apriori algorithm for solving the the frequent item sets problem [1]. This algorithm only considers a set to be a candidate frequent item set if all its subsets are already found to be frequent. In turn, the apriori algorithm is used to obtain association rules which establish predictions about how buying certain items implies buying other items.

More formally, let S be a set of items and let $\mathcal{B}$ be a nonempty finite set of baskets, each containing a set of items (i.e., a subset of S). Define $\mathcal{B}(X) = \{B \mid X \subseteq B \text{ and } B \in \mathcal{B}\}$. Observe that $\mathcal{B}(YZ) = \mathcal{B}(Y) \cap \mathcal{B}(Z)$ and that $\mathcal{B}(Y \cap Z) \supseteq \mathcal{B}(Y) \cup \mathcal{B}(Z)$. Define

$$\texttt{freq}(X) = \frac{|\mathcal{B}(X)|}{|\mathcal{B}|}.$$

Clearly, $\texttt{freq}(X) \geq \texttt{freq}(XY)$, whence `freq` is anti-isotone. It is also supermodular, since

$$\begin{aligned}|\mathcal{B}(Y \cap Z)| &\geq |\mathcal{B}(Y) \cup \mathcal{B}(Z)| \\ &= |\mathcal{B}(Y)| + |\mathcal{B}(Z)| - |\mathcal{B}(Y) \cap \mathcal{B}(Z)| \\ &= |\mathcal{B}(Y)| + |\mathcal{B}(Z)| - |\mathcal{B}(Y \cup Z)|.\end{aligned}$$

Hence, `freq` is an anti-isotone measure.

4 Additional inequalities for measures

In this section, we deduce other useful inequalities from the definitions of measures. Not surprisingly, some of these inequalities have already been used in applications for specific measures. As an example, consider rule 2 in Proposition 4, which states that, for each isotone measure $\mathcal{M}$, $\mathcal{M}(S) + \mathcal{M}(\emptyset) \leq \mathcal{M}(\overline{X}) + \mathcal{M}(X)$. When $\mathcal{M}$ is the Shannon entropy measure[8], this rule states that the entropy of S is smaller than the sum of the entropies of the components of a decomposition of S into two disjoint subsets [11]. As a second example, consider the analogon for anti-isotone measures of rule 1 in Proposition 4. When is $\mathcal{M}$ the `freq` measure of Section 3.3, this rule translates into the *apriori rule*, which states that if an item set is frequent then it subsets are also frequent.[9]

In addition to the bounds specified in Proposition 4, we consider bounds expressed in terms of certain average functions associated with measures. We will show that for an isotone measure (anti-isotone measure), that these associated average functions are increasing and concave (decreasing and concave, respectively).

Proposition 4. *Let $\mathcal{Y}$ be a nonempty set of subsets of S. If $\mathcal{M}$ is an isotone measure, then the following inequalities hold:*

1. $\max_{Y\in\mathcal{Y}} \mathcal{M}(XY) \leq \mathcal{M}(X \cup \bigcup \mathcal{Y})$;
2. $\min_{Y\in\mathcal{Y}} \mathcal{M}(X \cap Y) \geq \mathcal{M}(X \cap \bigcap \mathcal{Y})$;
3. $\mathcal{M}(X) + \mathcal{M}(\emptyset) \leq \mathcal{M}(X \cap \overline{Y}) + \mathcal{M}(X \cap Y)$;
4. $\mathcal{M}(S) + \mathcal{M}(\emptyset) \leq \mathcal{M}(\overline{X}) + \mathcal{M}(X)$;
5. $\mathcal{M}(X \cup \bigcup \mathcal{Y}) \leq \sum_{Y\in\mathcal{Y}} \mathcal{M}(XY) - (|\mathcal{Y}| - 1)\mathcal{M}(X)$;
6. $\mathcal{M}(X \cap \bigcap \mathcal{Y}) \geq \mathcal{M}(S) + \mathcal{M}(\emptyset) + (|\mathcal{Y}| - 1)\mathcal{M}(\overline{X}) - \sum_{Y\in\mathcal{Y}} \mathcal{M}(\overline{X \cap Y})$.

If $\mathcal{M}$ is an anti-isotone measure, then the above inequalities hold with the inequality signs reversed and with max *and* min *interchanged.*

Proof. Because of Proposition 2, it suffices to consider the case where $\mathcal{M}$ is an isotone measure.

Inequalities 1 and 2 follow from the isotonicity of $\mathcal{M}$.

Inequality 3 is established using submodularity, as follows:

$$\mathcal{M}(X) + \mathcal{M}(\emptyset) = \mathcal{M}(X \cap (Y\overline{Y})) + \mathcal{M}(\emptyset)$$

[8] Observe that, for the Shannon measure, $\mathcal{M}(\emptyset) = 0$.

[9] More advanced work on inequalities related to the `freq` measure occurs in [9]. The inequalities obtained there are derived from the inclusion-exclusion principle for counting finite sets.

$$
\begin{aligned}
&= \mathcal{M}((X \cap \overline{Y}) \cup (X \cap Y)) + \mathcal{M}((X \cap \overline{Y}) \cap (X \cap Y)) \\
&\leq \mathcal{M}(X \cap \overline{Y}) + \mathcal{M}(X \cap Y).
\end{aligned}
$$

Inequality 4 is a special case of inequality 3 for $X = S$.

Inequality 5 can be proved by induction on $|\mathcal{Y}|$. For $|\mathcal{Y}| = 1$, the inequality holds trivially. This is the basis for our induction. Let $k \geq 2$, and assume as induction hypothesis that inequality 5 holds for $|\mathcal{Y}| = k - 1$. We now prove inequality 5 for $|\mathcal{Y}| = k$. Let $Y \in \mathcal{Y}$. By the submodularity of $\mathcal{M}$,

$$
\begin{aligned}
\mathcal{M}(X \cup \bigcup \mathcal{Y}) = \mathcal{M}(XY \cup \; X \bigcup(\mathcal{Y} - \{Y\})) &\leq \mathcal{M}(XY) + \mathcal{M}(X \bigcup(\mathcal{Y} - \{Y\})) \\
&\quad - \mathcal{M}(XY \cap \; X \bigcup(\mathcal{Y} - \{Y\})) \\
&\leq \mathcal{M}(XY) + \mathcal{M}(X \bigcup(\mathcal{Y} - \{Y\})) \\
&\quad - \mathcal{M}(X)
\end{aligned}
$$

Inequality 5 then follows, since, by the induction hypothesis,

$$
\begin{aligned}
\mathcal{M}(X \cup \bigcup(\mathcal{Y} - \{Z\})) &\leq \textstyle\sum_{Y \in \mathcal{Y} - \{Z\}} \mathcal{M}(XY) - (|\mathcal{Y} - \{Z\}| - 1)\mathcal{M}(X) \\
&\leq \textstyle\sum_{Y \in \mathcal{Y} - \{Z\}} \mathcal{M}(XY) - (|\mathcal{Y}| - 2)\mathcal{M}(X)
\end{aligned}
$$

Inequality (6) can be deduced easily from inequalities 4 and 5 using De Morgan's laws:

$$
\begin{aligned}
\mathcal{M}(X \cap \bigcap \mathcal{Y}) &= \mathcal{M}\left(\overline{\textstyle\bigcup_{Y \in \mathcal{Y}} \overline{X} \cup \overline{Y}}\right) \\
&\geq \mathcal{M}(S) + \mathcal{M}(\emptyset) - \mathcal{M}(\textstyle\bigcup_{Y \in \mathcal{Y}} \overline{X} \cup \overline{Y}) \\
&\geq \mathcal{M}(S) + \mathcal{M}(\emptyset) - \textstyle\sum_{Y \in \mathcal{Y}} \mathcal{M}(\overline{X} \cup \overline{Y}) + (|\mathcal{Y}| - 1)\mathcal{M}(\overline{X}) \\
&= \mathcal{M}(S) + \mathcal{M}(\emptyset) + (|\mathcal{Y}| - 1)\mathcal{M}(\overline{X}) - \textstyle\sum_{Y \in \mathcal{Y}} \mathcal{M}(\overline{X \cap Y}).
\end{aligned}
$$

It is useful to consider Proposition 4 as providing lower and upper bounds for the measures of finite unions and intersections. In the case of isotone measures (when $X = \emptyset$, $X = S$ for inequalities (5) and (6) respectively), the bounds are as follows:

$$
\begin{aligned}
&\max_{Y \in \mathcal{Y}} \mathcal{M}(Y) \leq \mathcal{M}(\bigcup \mathcal{Y}) \leq \textstyle\sum_{Y \in \mathcal{Y}} \mathcal{M}(Y) - (|\mathcal{Y}| - 1)\mathcal{M}(\emptyset) \\
&\min_{Y \in \mathcal{Y}} \mathcal{M}(Y) \geq \mathcal{M}(\bigcap \mathcal{Y}) \geq |\mathcal{Y}|\mathcal{M}(\emptyset) + \mathcal{M}(S) - \textstyle\sum_{Y \in \mathcal{Y}} \mathcal{M}(\overline{Y}).
\end{aligned}
$$

In the case of anti-isotone measures, lower bounds and upper bounds must be swapped.

As shown in Proposition 4, useful bounds on measures of sets can be given in terms of measures of certain associated sets. In the rest of this section, we continue on this theme by considering how sums and averages of certain families of measures of sets are related. Thereto, consider the following definition:

Definition 2. Let X be a subset of S and let $\mathcal{Y}$ be a set of subsets of S such that $S = X \cup \bigcup \mathcal{Y}$. Let $n = |\mathcal{Y}|$, and let $0 \leq k \leq n$. Define

$$\mathcal{S}_k(X, \mathcal{Y}) = \sum_{\mathcal{Z} \subseteq \mathcal{Y} \wedge |\mathcal{Z}| = k} \mathcal{M}(X \cup \bigcup \mathcal{Z});$$
$$\mathcal{A}_k(X, \mathcal{Y}) = \frac{1}{\binom{n}{k}} \mathcal{S}_k(X, \mathcal{Y}).$$

Notice that $\mathcal{A}_0(X, \mathcal{Y}) = \mathcal{S}_0(X, \mathcal{Y}) = \mathcal{M}(X)$ and $\mathcal{A}_n(X, \mathcal{Y}) = \mathcal{S}_n(X, \mathcal{Y}) = \mathcal{M}(S)$.

We are mainly interested in $\mathcal{A}_k(X, \mathcal{Y})$, which is the average value of $\mathcal{M}(X \cup \bigcup \mathcal{Z})$ over all relevant choices of $\mathcal{Z}$, and, more in particular, how this quantity behaves as a function of k. In the next propositions, we show that, for fixed values of X and $\mathcal{Y}$, $\mathcal{A}_{.}(X, \mathcal{Y})$ is increasing (decreasing) when $\mathcal{M}$ is an isotone (anti-isotone) weak measure, and concave (convex) when $\mathcal{M}$ is an isotone (anti-isotone) measure.

Proposition 5. *Let $\mathcal{M}$ be an isotone weak measure, let $X \subseteq S$, and let $\mathcal{Y}$ be a set of subsets of S such that $S = X \cup \bigcup \mathcal{Y}$. Let $n = |\mathcal{Y}|$. Then $\mathcal{A}_{.}(X, \mathcal{Y})$ is an increasing function, i.e., for each k, $0 \leq k < n$,*

$$\mathcal{A}_k(X, \mathcal{Y}) \leq \mathcal{A}_{k+1}(X, \mathcal{Y}).$$

If $\mathcal{M}$ is an anti-isotone weak measure, then $\mathcal{A}_{.}(X, \mathcal{Y})$ is a decreasing function.

Proof. As usual, we only consider the case that $\mathcal{M}$ is isotone. Proposition 5 holds trivially if $n = 0$. When $n \geq 1$, Proposition 5 follows if we can prove that, for each k, $0 \leq k < n$,

$$\binom{n}{k+1} \mathcal{S}_k(X, \mathcal{Y}) \leq \binom{n}{k} \mathcal{S}_{k+1}(X, \mathcal{Y}),$$

or equivalently,

$$(n - k)\mathcal{S}_k(X, \mathcal{Y}) \leq (k + 1)\mathcal{S}_{k+1}(X, \mathcal{Y}).$$

For each $\mathcal{Z} \subseteq \mathcal{Y}$ with $|\mathcal{Z}| = k$, and for each $Y \in \mathcal{Y} - \mathcal{Z}$, we have that

$$\mathcal{M}(X \cup \bigcup \mathcal{Z}) \leq \mathcal{M}(X \cup \bigcup \mathcal{Z} \cup Y),$$

by the isotonicity of $\mathcal{M}$. Therefore,

$$(n-k)\mathcal{M}(X \cup \bigcup \mathcal{Z}) \leq \sum_{Y \in \mathcal{Y} - \mathcal{Z}} \mathcal{M}(X \cup \bigcup \mathcal{Z} \cup Y).$$

Summing the lefthand and righthand sides of this inequality over all $\mathcal{Z} \subseteq \mathcal{Y}$ with $|\mathcal{Z}| = k$, yields

$$(n-k)\mathcal{S}_k(X, \mathcal{Y}) \leq \sum_{\mathcal{Z} \subseteq \mathcal{Y} \wedge |\mathcal{Z}|=k} \sum_{Y \in \mathcal{Y} - \mathcal{Z}} \mathcal{M}(X \cup \bigcup \mathcal{Z} \cup Y) = (k+1)\mathcal{S}_{k+1}(X, \mathcal{Y}),$$

since there are $k+1$ different ways to write a subset of $\mathcal{Y}$ of size k as $\mathcal{Z} \cup Y$, with $\mathcal{Z} = k$ and $Y \in \mathcal{Y} - \mathcal{Z}$.

Proposition 6. *Let $\mathcal{M}$ be an isotone measure, let $X \subseteq S$, and let $\mathcal{Y}$ be a set of subsets of S such that $S = X \cup \bigcup \mathcal{Y}$. Let $n = |\mathcal{Y}|$. Then $\mathcal{A}_{.}(X, \mathcal{Y})$ is a concave function, i.e., for each k, $0 < k < n$,*

$$\mathcal{A}_k(X, \mathcal{Y}) \geq \frac{\mathcal{A}_{k-1}(X, \mathcal{Y}) + \mathcal{A}_{k+1}(X, \mathcal{Y})}{2}.$$

If $\mathcal{M}$ is an anti-isotone measure, then $\mathcal{A}_{.}(X, \mathcal{Y})$ is a convex function.

Proof. As usual, we only consider the case that $\mathcal{M}$ is isotone. The proposition is trivially true for $n \leq 1$. Therefore, assume $n \geq 2$. Let $\mathcal{Z} \subset \mathcal{Y}$ with $|\mathcal{Z}| = k-1$, and let Y_1 and Y_2 be two different elements of $\mathcal{Y} - \mathcal{Z}$. These elements exist, since $n \geq 2$ and $|\mathcal{Z}| \leq n-2$. By the submodularity of $\mathcal{M}$, it follows that

$$\mathcal{M}(X \cup \bigcup \mathcal{Z} \cup Y_1) + \mathcal{M}(X \cup \bigcup \mathcal{Z} \cup Y_2) \geq \mathcal{M}(X \cup \bigcup \mathcal{Z}) + \mathcal{M}(X \cup \bigcup \mathcal{Z} \cup Y_1 \cup Y_2).$$

As in the proof of Proposition 5, it follows that

$$\sum_{Y_1 \in \mathcal{Y} - \mathcal{Z}} \sum_{\mathcal{Z} \subseteq \mathcal{Y} \wedge |\mathcal{Z}|=k-1} \mathcal{M}(X \cup \bigcup \mathcal{Z} \cup Y_1) = k\mathcal{S}_k(X, \mathcal{Y}).$$

Since, given $\mathcal{Z}$ and Y_1, there remain $n-k$ possible choices for Y_2, we furthermore have that

$$\sum_{Y_1, Y_2 \in \mathcal{Y} - \mathcal{Z} \wedge Y_1 \neq Y_2} \sum_{\mathcal{Z} \subseteq \mathcal{Y} \wedge |\mathcal{Z}|=k-1} \mathcal{M}(X \cup \bigcup \mathcal{Z} \cup Y_1) = (n-k)k\mathcal{S}_k(X, \mathcal{Y}).$$

By symmetry, summing the entire lefthand side of the inequality above over all possible combinations of $\mathcal{Z}$, Y_1, and Y_2 yields the result $2(n-k)k\mathcal{S}_k(X,\mathcal{Y})$.

With regard to the righthand side of the inequality, similar arguments as the ones used for the lefthand side yield that

$$\sum_{Y_1,Y_2\in\mathcal{Y}-\mathcal{Z}\,\wedge\,Y_1\neq Y_2}\ \sum_{\mathcal{Z}\subseteq\mathcal{Y}\,\wedge\,|\mathcal{Z}|=k-1} \mathcal{M}(X\cup\bigcup\mathcal{Z}) = (n-k+1)(n-k)\mathcal{S}_{k-1}(X,\mathcal{Y}),$$

and that

$$\sum_{Y_1,Y_2\in\mathcal{Y}-\mathcal{Z}\,\wedge\,Y_1\neq Y_2}\ \sum_{\mathcal{Z}\subseteq\mathcal{Y}\,\wedge\,|\mathcal{Z}|=k-1} \mathcal{M}(X\cup\bigcup\mathcal{Z}\cup Y_1\cup Y_2) = (k+1)k\mathcal{S}_{k+1}(X,\mathcal{Y}).$$

Hence, summing the righthand side of the inequality over all possible $\mathcal{Z}$, Y_1, and Y_2 yields the result $(n-k+1)(n-k)\mathcal{S}_{k-1}(X,\mathcal{Y}) + (k+1)k\mathcal{S}_{k+1}(X,\mathcal{Y})$.

Thus, we obtain the inequality

$$2k(n-k)\mathcal{S}_k(X,\mathcal{Y}) \geq (n-k+1)(n-k)\mathcal{S}_{k-1}(X,\mathcal{Y})+(k+1)k\mathcal{S}_{k+1}(X,\mathcal{Y}).$$

Rewriting this inequality in terms of the averages yields the desired inequality.

We can use Proposition 5 and Proposition 6 to obtain lower and upper bounds on measure averages.

Proposition 7. *Let $\mathcal{M}$ be an isotone measure, let $X \subseteq S$, and let $\mathcal{Y}$ be a set of subsets of S such that $S = X \cup \bigcup\mathcal{Y}$. Let $n = |\mathcal{Y}|$. Then, for each k, $0 < k < n$,*

$$\mathcal{A}_k(X,\mathcal{Y}) \leq \mathcal{A}_{k+1}(X,\mathcal{Y}) \leq 2\mathcal{A}_k(X,\mathcal{Y}) - \mathcal{A}_{k-1}(X,\mathcal{Y}).$$

It follows from Proposition 7 that, if there exists k, $0 < k \leq n$, such that $\mathcal{A}_{k-1}(X,\mathcal{Y}) = \mathcal{A}_k(X,\mathcal{Y})$, then $\mathcal{A}_k(X,\mathcal{Y}) = \mathcal{A}_{k+m}(X,\mathcal{Y})$ for all m such that $k+m \leq n$, i.e., as soon as the function $\mathcal{A}_{.}(X,\mathcal{Y})$ stops increasing (decreasing) *strictly*, it becomes constant.

5 Measure differentials

5.1 Finite differences and differentials

A natural problem that arises for measures as for other functions is to calculate their rate of change and the rate of change of their rate of change. For functions over continuous domains, these quantities are represented by the first and second derivatives, respectively. In our framework for measures, we are dealing with discrete, set-based functions, and thus reasoning about derivatives must be done by using the methods of finite differences and finite difference equations [17].

Definition 3. Let $f : 2^S \to \mathbf{R}$ be a (finite) set-based function into the reals. Then the *first finite difference* of f is the function $\Delta f : 2^S \times 2^S \to \mathbf{R}$ defined by

$$\Delta f(X, Y) = f(XY) - f(X).$$

For a fixed value of Y, the function $\Delta f(\,.\,, Y) : 2^S \to \mathbf{R} : V \mapsto \Delta f(V, Y)$ is again a (finite) set-based function into the reals, of which we can take again the first finite difference. This is then the second finite difference of f.

Definition 4. Let $f : 2^S \to \mathbf{R}$ be a (finite) set-based function into the reals. Then the *second finite difference* of f is the function $\Delta^2 f : 2^S \times 2^S \times 2^S \to \mathbf{R}$ defined by

$$\Delta^2 f(X, Y, Z) = \Delta\Delta f(\,.\,, Y)(X, Z).$$

Expanding the above equation yields

$$\begin{aligned}\Delta^2 f(X, Y, Z) &= \Delta\Delta f(\,.\,, Y)(X, Z)\\ &= \Delta f(XZ, Y) - \Delta f(X, Y)\\ &= f(XYZ) - f(XZ) - f(XZ) + f(X).\end{aligned}$$

Notice that the above expression is invariant under swapping Y and Z.

The following proposition is now straightforward.

Proposition 8. *1. The first finite difference of a weak measure is nonnegative if it is isotone and nonpositive if it is anti-isotone.*

2. *The first finite difference of a measure is nonnegative if it is isotone and nonpositive if it is anti-isotone. The second finite difference of a measure is nonpositive if it is isotone and nonnegative if it is anti-isotone.*

Thus, observe that the signature of an isotone (anti-isotone) weak measure is that of an increasing (a decreasing) function, and that of a isotone (an anti-isotone) measure is that of an increasing, concave (a decreasing, convex) function.

We will use the concept of finite differences to reason about measures. Intuitively, these differences should calculate, when applied to sets, a value of how dependent/independent the measure of a set is with respect to a change. As an example, the first difference for the `max` measure would calculate the degree of change of $\mathtt{max}(X)$ with respect to $\mathtt{max}(XY)$. A small value for this difference would imply that the measure of set X is largely independent of a change of X by Y. In particular, when this difference is zero, i.e., $\mathtt{max}(XY) = \mathtt{max}(X)$, this can be interpreted as the value of the $\mathtt{max}(X)$ reaching an extreme point with respect to the maximum of the set Y. In fact, the condition $\mathtt{max}(XY) = \mathtt{max}(X)$ is equivalent to $\mathtt{max}(X) \geq \mathtt{max}(Y)$.

We wish to present properties in the form of inequality rules for the first and second finite differences of (weak) measures. As may already be expected, this is a somewhat awkward, because we always have to distinguish between the isotone and the anti-isotone case. The following proposition serves as an example for this awkwardness.

Proposition 9. *1. If $\mathcal{M}$ is an isotone (weak) measure, then*

$$\Delta\mathcal{M}(X, Y_1) \geq \Delta\mathcal{M}(X, Y_2) \quad \textit{if} \quad Y_2 \subseteq Y_1;$$

If $\mathcal{M}$ is an anti-isotone (weak) measure, then the above inequality holds with the inequality sign reversed.

2. *If $\mathcal{M}$ is an isotone measure, then*

$$\Delta^2\mathcal{M}(X, Y, Z_1) \leq \Delta^2\mathcal{M}(X, Y, Z_2) \quad \textit{if} \quad Z_2 \subseteq Z_1;$$

If $\mathcal{M}$ is an anti-isotone measure, then the above inequality holds with the inequality sign reversed.

Proof. We only consider the isotone case; the anti-isotone case is completely analogous.

1. We prove the inequality by expansion:

$$\Delta\mathcal{M}(X,Y_1) - \Delta\mathcal{M}(X,Y_2) = \mathcal{M}(XY_1) - \mathcal{M}(X) - \mathcal{M}(XY_2) + \mathcal{M}(X)$$
$$= \mathcal{M}(XY_1) - \mathcal{M}(XY_2) \geq 0,$$

   by isotonicity.
2. We prove the inequality by expansion. Let $U = XZ_2$.

$$\begin{aligned}
&\Delta^2\mathcal{M}(X,Y,Z_2) - \Delta^2\mathcal{M}(X,Y,Z_1)\\
&\quad= \mathcal{M}(XYZ_2) - \mathcal{M}(XY) - \mathcal{M}(XZ_2) + \mathcal{M}(X)\\
&\qquad\qquad - \mathcal{M}(XYZ_1) + \mathcal{M}(XY) + \mathcal{M}(XZ_1) - \mathcal{M}(X)\\
&\quad= \mathcal{M}(XYZ_2) + M(XZ_1) - \mathcal{M}(XYZ_1) - \mathcal{M}(XZ_2)\\
&\quad= \mathcal{M}(UY) + M(U(Z_1 - Z_2)) - \mathcal{M}(UY(Z_1 - Z_2)) - \mathcal{M}(U) \geq 0,
\end{aligned}$$

   by Proposition 3.

As Proposition 9 clearly illustrates, rules formulated in terms of finite differences have the awkwardness that the inequality sign is different for isotone and anti-isotone (weak) measures. Therefore, we introduce first and second *finite differentials*. These are the absolute values of the related finite differences. Their introduction has the advantage that subsequent rules for them are the same for both isotone and anti-isotone (weak) measures. The notation introduced for first and second finite differences in Definition 5 below reflects an analogy with database dependencies which will be made explicit in Section 6.

Definition 5. Let $f : 2^S \to \mathbf{R}$ be a (finite) set-based function into the reals.

1. The *first finite differential* of f is the function $f' : 2^S \times 2^S \to \mathbf{R}$ defined by
$$f'(V,W) = |\Delta f(V,W)|,$$
   which will be denoted as $f'(V \to W)$.
2. The *second finite differential* of f is the function $f'' : 2^S \times 2^S \times 2^S \to \mathbf{R}$ defined by
$$f''(V,W_1,W_2) = |\Delta^2 f(V,W_1,W_2)|,$$
   which will be denoted as $f''(V \twoheadrightarrow W_1|W_2)$.

As an example to illustrate the advantage of finite differentials over finite differences, we restate Proposition 9 in terms of finite differentials.

Proposition 10. *1. If $\mathcal{M}$ is a (weak) measure, then*

$$\mathcal{M}'(X \to Y_1) \geq \mathcal{M}'(X \to Y_2) \quad \textit{if} \quad Y_2 \subseteq Y_1.$$

2. If $\mathcal{M}$ is a measure, then

$$\mathcal{M}''(X \twoheadrightarrow Y|Z_1) \geq \mathcal{M}''(X \twoheadrightarrow Y|Z_2) \quad \textit{if} \quad Z_2 \subseteq Z_1.$$

Clearly, Proposition 10 is much more elegant and workable than Proposition 9. In the remaining subsections, we present rules for first differentials, rules for second differential, and mixed rules, respectively.

5.2 Properties of first differentials

In this subsection, we specify equalities and inequalities for the first differentials of (weak) measures.

Proposition 11. *Let $\mathcal{M}$ be a (weak) measure. Then the first finite differential of $\mathcal{M}$ satisfies the following inequalities:*

$$\begin{array}{ll}
\mathcal{M}'(X \to Y) = 0 \quad \textit{if} \quad Y \subseteq X & \text{(triviality);} \\
\mathcal{M}'(X \to Y) \geq \mathcal{M}'(XU \to VY) & \textit{if} \quad V \subseteq U \subseteq XY \\
 & \text{(weak augmentation);} \\
\mathcal{M}'(X \to Y) + \mathcal{M}'(Y \to Z) \geq \mathcal{M}'(X \to Z) & \textit{if} \quad X \subseteq Z \\
 & \text{(weak transitivity).}
\end{array}$$

Proof. As usual, we only consider the case that $\mathcal{M}$ is isotone. Triviality follows, since if $X \subseteq Y$, then

$$\mathcal{M}'(X \to Y) = \mathcal{M}(XY) - \mathcal{M}(X) = \mathcal{M}(X) - \mathcal{M}(X) = 0.$$

Weak augmentation follows, since, by the isotonicity of $\mathcal{M}$,

$$\begin{aligned}
\mathcal{M}'(XU \to VY) &= \mathcal{M}(XUVY) - \mathcal{M}(XU) \\
&= \mathcal{M}(XY) - \mathcal{M}(XU) \\
&\leq \mathcal{M}(XY) - \mathcal{M}(X) \\
&= \mathcal{M}'(X \to Y)
\end{aligned}$$

Weak transitivity follows, since $\mathcal{M}'(X \to Y) + \mathcal{M}'(Y \to Z) - \mathcal{M}'(X \to Z)$ is equal to

$$\begin{aligned}
&\mathcal{M}(XY) - \mathcal{M}(X) + \mathcal{M}(YZ) - \mathcal{M}(Y) - \mathcal{M}(XZ) + \mathcal{M}(X) \\
&\quad = \mathcal{M}(XY) + \mathcal{M}(YZ) - \mathcal{M}(Y) - \mathcal{M}(XZ) \\
&\quad = \mathcal{M}(XY) - \mathcal{M}(Y) + \mathcal{M}(YZ) - \mathcal{M}(Z) \geq 0,
\end{aligned}$$

by isotonicity.

For measures, Proposition 11 can be strengthened, as follows.

Proposition 12. *Let $\mathcal{M}$ be a measure. Then the first finite differential of $\mathcal{M}$ satisfies the following inequalities:*

$$\begin{array}{ll}
\mathcal{M}'(X \to Y) = 0 \quad \textit{if} \;\; Y \subseteq X & \text{(triviality);} \\
\mathcal{M}'(X \to Y) \geq \mathcal{M}'(XU \to VY) \quad \textit{if} \;\; V \subseteq U & \text{(augmentation);} \\
\mathcal{M}'(X \to Y) + \mathcal{M}'(Y \to Z) \geq \mathcal{M}'(X \to Z) & \text{(transitivity).}
\end{array}$$

Proof. As usual, we only consider the case that $\mathcal{M}$ is isotone. Triviality follows from Proposition 11. Augmentation follows from

$$\mathcal{M}'(XU \to VY) = \mathcal{M}(XUVY) - \mathcal{M}(XU) \leq \mathcal{M}(XY) - \mathcal{M}(X) = \mathcal{M}'(X \to Y),$$

where Proposition 3 with Z set to U is used to justify the inequality. Transitivity follows, since by Proposition 3 and isotonicity,

$$\begin{aligned}
\mathcal{M}'(X \to Y) + \mathcal{M}'(Y \to Z) &= \mathcal{M}(XY) - \mathcal{M}(X) + \mathcal{M}(YZ) - \mathcal{M}(Y) \\
&\geq \mathcal{M}(XYZ) - \mathcal{M}(X) \\
&\geq \mathcal{M}(XZ) - \mathcal{M}(X) \\
&= \mathcal{M}'(X \to Z).
\end{aligned}$$

Example 1. An example of an isotone weak measure that is not a measure is `max2`, which returns the second largest element of a set. This measure does not satisfy augmentation and transitivity (Proposition 12). For example, if $X = \{2, 3\}$, $Y = \{1, 6\}$, and $U = \{3, 5\}$ then augmentation does not hold. However, weak augmentation (Proposition 11) holds when $U \subseteq XY$ (e.g., $X = \{2, 3\}$, $Y = \{1, 6\}$, and $U = \{2, 6\}$). Similarly `max2` is not transitive (e.g., $X = \{1, 6\}$, $Y = \{2, 3\}$, and $Z = \{1, 2, 5\}$). However, weak transitivity holds when $X \subseteq Z$ (e.g., $X = \{1, 2\}$, $Y = \{2, 3\}$, and $Z = \{1, 2, 5\}$).

5.3 Properties of second differentials

Proposition 13. *Let $\mathcal{M}$ be a measure. Then the second finite differential of $\mathcal{M}$ satisfies the following inequalities:*

$$\begin{aligned}
&\mathcal{M}''(X \twoheadrightarrow Y|Z) = 0 \quad \textit{if} \;\; Y \subseteq Z && \text{(triviality);}\\
&\mathcal{M}''(X \twoheadrightarrow Y|Z) = \mathcal{M}''(X \twoheadrightarrow Z|Y) && \text{(symmetry);}\\
&\mathcal{M}''(X \twoheadrightarrow Y|Z) \geq \mathcal{M}''(XU \twoheadrightarrow YV|Z) \quad \textit{if} \;\; V \subseteq U \subseteq XYZ &&\\
& && \text{(weak augmentation);}\\
&\mathcal{M}''(X \twoheadrightarrow Y|ZU) + \mathcal{M}''(Y \twoheadrightarrow Z|XU) \geq \mathcal{M}''(X \twoheadrightarrow Z - Y|YU) &&\\
& && \text{(weak transitivity).}
\end{aligned}$$

Proof. As usual, we only consider the case that $\mathcal{M}$ is isotone. Triviality and complementation follow directly from Definition 5.

To prove weak augmentation we first prove the seemingly weaker rule below:

$$\mathcal{M}''(X \twoheadrightarrow Y|Z) \geq \mathcal{M}''(XU \twoheadrightarrow YV|Z) \text{ if } V \subseteq U \subseteq XY \textit{ or } V \subseteq U \subseteq XZ.$$

By expanding, we find that

$$\begin{aligned}
\mathcal{M}''(X &\twoheadrightarrow Y|Z) - \mathcal{M}''(XU \twoheadrightarrow YV|Z)\\
&= \mathcal{M}(XY) + \mathcal{M}(XZ) - \mathcal{M}(X) - \mathcal{M}(XYZ) - \mathcal{M}(XUYV)\\
&\qquad - \mathcal{M}(XUZ) + \mathcal{M}(XU) + \mathcal{M}(XUYVZ)\\
&= \mathcal{M}(XY) + \mathcal{M}(XZ) - \mathcal{M}(X) - \mathcal{M}(XYZ) - \mathcal{M}(XUY)\\
&\qquad - \mathcal{M}(XUZ) + \mathcal{M}(XU) + \mathcal{M}(XUYZ),
\end{aligned}$$

taking into account that $V \subseteq U$. The last expression is invariant under interchanging Y and Z. We may therefore assume, without loss of generality, that $U \subseteq XY$. Thus,

$$\begin{aligned}
\mathcal{M}(XY) &+ \mathcal{M}(XZ) - \mathcal{M}(X) - \mathcal{M}(XYZ) - \mathcal{M}(XUY)\\
&\qquad - \mathcal{M}(XUZ) + \mathcal{M}(XU) + \mathcal{M}(XUYZ)\\
&= \mathcal{M}(XY) + \mathcal{M}(XZ) - \mathcal{M}(X) - \mathcal{M}(XYZ) - \mathcal{M}(XY)\\
&\qquad - \mathcal{M}(XUZ) + \mathcal{M}(XU) + \mathcal{M}(XYZ)\\
&= \mathcal{M}(XZ) - \mathcal{M}(X) - \mathcal{M}(XUZ) + \mathcal{M}(XU),
\end{aligned}$$

which is nonnegative by Proposition 3. We now bootstrap our intermediate result to what we actually want to prove. Thus, assume that $V \subseteq U \subseteq XYZ$, and let $V_Y = V \cap XY$, $U_Y = U \cap XY$, $V_Z = V \cap XZ$, and $U_Z = U \cap XZ$. Clearly, $V_Y V_Z = V$ and $U_Y U_Z = U$. Since $V_Y \subseteq U_Y \subseteq XY$, the intermediate result above guarantees that

$$\mathcal{M}''(X \twoheadrightarrow Y|Z) \geq \mathcal{M}''(XU_Y \twoheadrightarrow YV_Y|Z).$$

Since $V_Z \subseteq U_Z \subseteq XU_Y Z$, our intermediate result also guarantees that

$$\mathcal{M}''(XU_Y \twoheadrightarrow YV_Y|Z) \geq \mathcal{M}''(XU_Y U_Z \twoheadrightarrow YV_Y V_Z|Z).$$

Combined both inequalities above yield the desired result.

Weak transitivity can be proved as follows.

$$\begin{aligned}
&\mathcal{M}''(X \twoheadrightarrow Y|ZU) + \mathcal{M}''(Y \twoheadrightarrow Z|XU) - \mathcal{M}''(X \twoheadrightarrow Z - Y|YU) \\
&\quad \geq \mathcal{M}''(X \twoheadrightarrow Y|ZU) + \mathcal{M}''(Y \twoheadrightarrow Z|XU) - \mathcal{M}''(X \twoheadrightarrow Z|YU) \\
&\quad = \mathcal{M}(XY) + \mathcal{M}(XZU) - \mathcal{M}(X) - \mathcal{M}(XYZU) + \mathcal{M}(YZ) \\
&\qquad\qquad + \mathcal{M}(YXU) - \mathcal{M}(Y) - \mathcal{M}(XYZU) - \mathcal{M}(XZ) \\
&\qquad\qquad - \mathcal{M}(XYU) + \mathcal{M}(X) + \mathcal{M}(XYZU) \\
&\quad = \mathcal{M}(XY) + \mathcal{M}(XZU) + \mathcal{M}(YZ) - \mathcal{M}(Y) - \mathcal{M}(XYZU) \\
&\qquad\qquad - \mathcal{M}(XZ).
\end{aligned}$$

By applying Proposition 3 to the first and third terms of the above expression, we find that

$$\begin{aligned}
&\mathcal{M}(XY) + \mathcal{M}(XZU) + \mathcal{M}(YZ) - \mathcal{M}(Y) - \mathcal{M}(XYZU) - \mathcal{M}(XZ) \\
&\quad \geq \mathcal{M}(Y) + \mathcal{M}(XZU) + \mathcal{M}(XYZ) - \mathcal{M}(Y) - \mathcal{M}(XYZU) - \mathcal{M}(XZ) \\
&\quad = \mathcal{M}(XZU) + \mathcal{M}(XYZ) - \mathcal{M}(XYZU) - M(XZ) \geq 0,
\end{aligned}$$

again by Proposition 3.

An important special case of the second finite differential $\mathcal{M}''(X \twoheadrightarrow Y|Z)$ occurs when $Z = S - XY$, where S is the universe over which $\mathcal{M}$ is defined. We therefore introduce the notation $\mathcal{M}''(X \twoheadrightarrow Y)$ as shorthand for $\mathcal{M}''(X \twoheadrightarrow Y|S - XY)$. The following proposition now follows easily.

Proposition 14. *Let $\mathcal{M}$ be a measure. Then the second finite differential of $\mathcal{M}$ satisfies the following inequalities:*

$$\begin{array}{ll}
\mathcal{M}''(X \twoheadrightarrow Y) = 0 \ \ \textit{if} \ \ Y \subseteq X & \text{(triviality);} \\
\mathcal{M}''(X \twoheadrightarrow Y) = \mathcal{M}''(X \twoheadrightarrow S - XY) & \text{(complementation);} \\
\mathcal{M}''(X \twoheadrightarrow Y) \geq \mathcal{M}''(XU \twoheadrightarrow YV) \ \ \textit{if} \ \ V \subseteq U & \text{(augmentation);} \\
\mathcal{M}''(X \twoheadrightarrow Y) + \mathcal{M}''(Y \twoheadrightarrow Z) \geq \mathcal{M}''(X \twoheadrightarrow Z - Y) & \text{(transitivity).}
\end{array}$$

Proof. We only prove transitivity, since the other rules immediate follow from Proposition 13. Let $U = S - XYZ$. By Proposition 10, $\mathcal{M}''(Y \twoheadrightarrow Z|XU) \geq \mathcal{M}''(Y \twoheadrightarrow Z - Y|XU)$. Therefore, $\mathcal{M}''(X \twoheadrightarrow Y)+\mathcal{M}''(Y \twoheadrightarrow Z) \geq \mathcal{M}''(X \twoheadrightarrow Y|(Z-Y)U)+\mathcal{M}''(Y \twoheadrightarrow Z-Y|XU)$. By weak transitivity, this sum is at least $\mathcal{M}''(X \twoheadrightarrow Z - Y)$.

5.4 Mixed inequalities

In the final proposition of this section, we state mixed inequalities between first and second differentials.

Proposition 15. *Let $\mathcal{M}$ be a measure. Then the first and second finite differentials of $\mathcal{M}$ satisfy the following inequalities:*

$$\begin{array}{ll} \mathcal{M}'(X \to Y) \geq \mathcal{M}''(X \twoheadrightarrow Y|Z) & \text{(replication);} \\ \mathcal{M}''(X \twoheadrightarrow Y|Z) + \mathcal{M}'(Y \to Z) \geq \mathcal{M}'(X \to Z - Y) & \text{(coalescence).} \end{array}$$

Proof. As usual, we only consider the case that $\mathcal{M}$ is isotone. Replication follows by straightforwardly expanding $\mathcal{M}'(X \to Y)-\mathcal{M}''(X \twoheadrightarrow Y|Z)$ and applying isotonicity.

Coalescence can be proved as follows.

$$\begin{aligned} &\mathcal{M}''(X \twoheadrightarrow Y|Z) + \mathcal{M}'(Y \to Z) - \mathcal{M}'(X \to Z - Y) \\ &\quad \geq \mathcal{M}''(X \twoheadrightarrow Y|Z) + \mathcal{M}'(Y \to Z) - \mathcal{M}'(X \to Z) \\ &\quad = \mathcal{M}(XY) + \mathcal{M}(XZ) - \mathcal{M}(X) - \mathcal{M}(XYZ) + \mathcal{M}(YZ) \\ &\qquad\qquad - \mathcal{M}(Y) - \mathcal{M}(XZ) + \mathcal{M}(X) \\ &\quad = \mathcal{M}(XY) - \mathcal{M}(XYZ) + \mathcal{M}(YZ) - \mathcal{M}(Y) \geq 0, \end{aligned}$$

by Proposition 3.

Corollary 1. *Let $\mathcal{M}$ be a measure. Then the first and second finite differentials of $\mathcal{M}$ satisfy the following inequalities:*

$$\begin{array}{ll} \mathcal{M}'(X \to Y) \geq \mathcal{M}''(X \twoheadrightarrow Y) & \text{(replication);} \\ \mathcal{M}''(X \twoheadrightarrow Y) + \mathcal{M}'(Y \to Z) \geq \mathcal{M}'(X \to Z - Y) & \text{(coalescence).} \end{array}$$

Proof. The replication rule in Corollary 1 is a straightforward translation of the replication rule in Proposition 15. To prove the coalescence rule of Corollary 1, first observe that $\mathcal{M}''(X \twoheadrightarrow Y) \geq \mathcal{M}''(X \twoheadrightarrow Y|Z-Y)$ and $\mathcal{M}'(Y \to Z) \geq \mathcal{M}'(Y \to Z-Y)$, by Proposition 10. The coalescence rule of Corollary 1 now follows straightforwardly from the coalescence rule of Proposition 15.

6 Measure constraints

In this section, we consider the situations in which the first and second differentials are zero. This leads us to introduce *primary* and *secondary constraints* and to derive inference rules for them.

Satisfaction of constraints by particular measures discussed in Section 3 yields conditions in terms of satisfaction of constraints in databases and data mining.

6.1 Primary constraints

Definition and inference rules

Definition 6. Let $\mathcal{M}$ be a (weak) measure. Then $\mathcal{M}$ satisfies the *primary constraint* $X \to Y$ if $\mathcal{M}'(X \to Y) = 0$.

Propositions 11 and 12 yield the following inference rules for primary constraints.

Proposition 16. *The following inference rules hold for primary constraints with respect to (weak) measures:*

$$\begin{array}{ll} X \to Y \quad \textit{if} \quad Y \subseteq X & \text{(triviality);} \\ X \to Y \ \textit{implies} \ XU \to YV \quad \textit{if} \quad V \subseteq U \subseteq XY & \text{(weak augmentation);} \\ X \to Y \ \textit{and} \ Y \to Z \ \textit{imply} \ X \to Z \quad \textit{if} \quad X \subseteq Z & \text{(weak transitivity).} \end{array}$$

With respect to measures, *these inference rules can be strengthened, as follows:*

$$\begin{array}{ll} X \to Y \quad \textit{if} \quad Y \subseteq X & \text{(triviality);} \\ X \to Y \ \textit{implies} \ XU \to YV \quad \textit{if} \quad V \subseteq U & \text{(augmentation);} \\ X \to Y \ \textit{and} \ Y \to Z \ \textit{imply} \ X \to Z & \text{(transitivity).} \end{array}$$

We now interpret primary constraints for various measures discussed in Section 3.

Databases—aggregate functions The use of aggregate functions in databases was described in Section 3.1. In particular, we showed that `count`, `sum`, `max`, and `min` are measures. We now interpret satisfaction of primary constraints by these measures.

1. The `sum` measure satisfies the primary constraint $X \to Y$ if and only if $Y \subseteq X$. It is readily seen that inclusion satisfies the triviality, augmentation, and transitivity rules.
2. The `count` measure satisfies the primary constraint $X \to Y$ if and only if $Y \subseteq X$ (very similar to `sum`).
3. The `max` measure satisfies the primary constraint $X \to Y$ if and only if $\mathtt{max}(X) \geq \mathtt{max}(Y)$. Again it can be readily verified that the latter constraint satisfies the triviality, augmentation, and transitivity rules.
4. The `min` measure satisfies the primary constraint $X \to Y$ if and only if $\mathtt{min}(X) \leq \mathtt{min}(Y)$. Again it can be readily verified that the latter constraint satisfies the triviality, augmentation, and transitivity rules.

Databases—data uniformity Shannon's entropy measure and the Gini index were described in Section 3.2. Primary measure constraints can be used to capture database constraints, as shown below.

1. Let S be a relation scheme, let X and Y be subsets of S, and let ϱ be a relation instance over S. Let p be a probability distribution over ϱ satisfying $p(t) \neq 0$ for all t in ϱ. The corresponding Gini index $\mathcal{G}$ satisfies the primary constraint $X \to Y$ if and only if ϱ satisfies the functional dependency $X \overset{\text{fd}}{\to} Y$.
   To see this, let $Z = XY$. To show that $\mathcal{G}(X) = \mathcal{G}(Z)$, we have to show that
   $$\sum_{x \in \pi_X(\varrho)} p_X^2(x) = \sum_{z \in \pi_Z(\varrho)} p_Z^2(z)$$
   Now, let $\varrho_x = \sigma_{X=x}(\varrho)$. From the calculations in Section 3.2, we recall that
   $$p_X^2(x) = \Big[\sum_{z \in \pi_Z(\varrho_x)} p_Z(z) \Big]^2 \geq \sum_{z \in \pi_Z(\varrho_x)} p_Z^2(z).$$
   By our additional assumption on the probability distribution p, equality holds if and only if $|\pi_Z(\varrho_x)| = 1$. We also showed in Section 3.2 that summing up left-hand and right-hand side of the above inequality yields
   $$\mathcal{G}(X) \geq \mathcal{G}(Z).$$

Hence, equality holds if and only if $|\pi_Z(\varrho_x)| = 1$ for all x in $\pi_X(\varrho)$. This is clearly equivalent with saying that ϱ satisfies the functional dependency $X \stackrel{\text{fd}}{\to} Y$.
Notice that, in this case, the triviality, augmentation, and transitivity rules correspond to the well-known reflexivity, augmentation, and transitivity rules for functional dependencies [4].

2. Under the same assumptions as above, the Shannon entropy measure $\mathcal{H}$ satisfies the primary constraint $X \to Y$ if and only if ϱ satisfies the functional dependency $X \stackrel{\text{fd}}{\to} Y$.
To see this, let $\varrho_x = \sigma_{X=x}(\varrho)$. Then $\mathcal{H}(XY) = \mathcal{H}(X)$ is equivalent to

$$\sum_{x\in\pi_X(\varrho)} p_X(x)\log p_X(x) - \sum_{x\in\pi_X(\varrho)} \textstyle\sum_{y\in\pi_Y(\varrho_x)} p_{XY}(xy)\log p_{XY}(xy) = 0.$$

Using that $p_X(x) = \sum_{y\in\pi_Y(\varrho_x)} p_{XY}(xy)$, the above equality can be rewritten as

$$\sum_{x\in\pi_X(\varrho)} \textstyle\sum_{y\in\pi_Y(\varrho_x)} p_{XY}(xy)(\log p_X(x) - \log p_{XY}(xy)) = 0.$$

Since each term in the sum is nonnegative, the last equality is equivalent to $p_X(x) = p_{XY}(xy)$, for all x in $\pi_X(\varrho)$ and for all y in $\pi_Y(\varrho_x)$. Clearly, this condition is equivalent to ϱ satisfying $X \stackrel{\text{fd}}{\to} Y$.

Data mining—frequent item sets The use of the `freq` measure to measure frequent item sets in data mining was described in Section 3.3 The `freq` measure satisfies the primary constraint $X \to Y$ if and only if $\texttt{freq}(X \cup Y) = \texttt{freq}(X)$ if and only if $\mathcal{B}(X \cup Y) = \mathcal{B}(X)$ if and only if there is a pure association rule from X to Y in $\mathcal{B}$ [1]. (A pure association rule is an association rule with confidence 100%.) In this context, the triviality, augmentation, and transitivity rules for primary constraints can be interpreted as properties and as inference rules of pure association rules.

6.2 Secondary constraints

Definition and inference rules

Definition 7. Let $\mathcal{M}$ be a measure. Then

1. $\mathcal{M}$ satisfies the *secondary constraint* $X \twoheadrightarrow Y|Z$ if $\mathcal{M}''(X \twoheadrightarrow Y|Z) = 0$;
2. $\mathcal{M}$ satisfies the *strong secondary constraint* $X \twoheadrightarrow Y$ if it satisfies the secondary constraint $X \twoheadrightarrow Y|S - Y$.

Propositions 13 and 14 yield the following inference rules.

Proposition 17. *The following inference rules hold for secondary constraints with respect to measures:*

$$
\begin{array}{ll}
X \twoheadrightarrow Y|Z \;\; \textit{if} \;\; Y \subseteq X & \text{(triviality);} \\
X \twoheadrightarrow Y|Z \textit{ implies } X \twoheadrightarrow Z|Y & \text{(symmetry);} \\
X \twoheadrightarrow Y|Z \textit{ implies } XU \twoheadrightarrow VY|Z \;\; \textit{if} \;\; V \subseteq U \subseteq XYZ & \\
 & \text{(weak augmentation);} \\
X \twoheadrightarrow Y|ZU \textit{ and } Y \twoheadrightarrow Z|XU \textit{ imply } X \twoheadrightarrow Z - Y|YU & \text{(weak transitivity).}
\end{array}
$$

For strong *secondary constraints, these inference rules can be strengthened, as follows:*

$$
\begin{array}{ll}
X \twoheadrightarrow Y \;\; \textit{if} \;\; Y \subseteq X & \text{(triviality);} \\
X \twoheadrightarrow Y \textit{ implies } X \twoheadrightarrow S - Y & \text{(complementation);} \\
X \twoheadrightarrow Y \textit{ implies } XU \twoheadrightarrow YV \;\; \textit{if} \;\; V \subseteq U & \text{(augmentation);} \\
X \twoheadrightarrow Y \textit{ and } Y \twoheadrightarrow Z \textit{ imply } X \twoheadrightarrow Z - Y & \text{(transitivity).}
\end{array}
$$

We also present some mixed inference rules for primary and secondary constraints.

Proposition 18. *The following inference rules hold for primary and secondary constraints with respect to measures:*

$$
\begin{array}{ll}
X \rightarrow Y \textit{ implies } X \twoheadrightarrow Y|Z & \text{(replication);} \\
X \twoheadrightarrow Y|Z \textit{ and } Y \rightarrow Z \textit{ imply } X \rightarrow Z - Y & \text{(coalescence).}
\end{array}
$$

For primary and strong *secondary constraints, these inference rules become as follows:*

$$
\begin{array}{ll}
X \rightarrow Y \textit{ implies } X \twoheadrightarrow Y & \text{(replication);} \\
X \twoheadrightarrow Y \textit{ and } Y \rightarrow Z \textit{ imply } X \rightarrow Z - Y & \text{(coalescence).}
\end{array}
$$

We now interpret secondary constraints for various measures discussed in Section 3.

Databases—aggregate functions We briefly revisit aggregate functions in databases and interpret satisfaction of secondary constraints by these measures, in particular for `count` and `max`.

1. Let X, Y, and Z be subsets of the set S. Then the `count` measure satisfies the secondary constraint $X \twoheadrightarrow Y|Z$ if and only if $\texttt{count}(XY) + \texttt{count}(XZ) = \texttt{count}(XYZ) + \texttt{count}(X)$, which is true if and only if $Y \cap Z \subseteq X$. This implies a notion of independence between Y and Z. For example, in the case where $X = \emptyset$, it follows that Y_1 and Y_2 must be disjoint.
2. Let X, Y, and Z be subsets of the set of nonnegative numbers S. Then the `max` measure satisfies the secondary constraint $X \twoheadrightarrow Y|Z$ if and only if $\texttt{max}(XY) + \texttt{max}(XZ) = \texttt{max}(XYZ) + \texttt{max}(X)$, which holds if and only if $\texttt{max}(X) \geq \texttt{max}(Y)$ or $\texttt{max}(X) \geq \texttt{max}(Z)$. Observe that, in the case where $X = \emptyset$, either Y or Z must be the empty set.

Databases—data uniformity Shannon's entropy measure and the Gini index were described in Section 3.2. Secondary measure constraints can also be used to capture database constraints, as shown below.

1. Let S be a relation scheme, let X and Y be subsets of S, and let ϱ be a nonempty relation instance over S. Let p be a probability distribution over ϱ satisfying $p(t) \neq 0$ for all t in ϱ. The corresponding Gini index $\mathcal{G}$ satisfies the secondary constraint $X \twoheadrightarrow Y$ if and only if, for all x in $\pi_X(\varrho)$, $\varrho_x = \sigma_{X=x}(\varrho)$ satisfies either the functional dependency $X \stackrel{\text{fd}}{\rightarrow} Y$ or the functional dependency $X \stackrel{\text{fd}}{\rightarrow} S - XY$.
   To see this, we first assume that X and Y are disjoint.
   Observe that $\mathcal{G}$ satisfies $X \twoheadrightarrow Y$ if and only if $\mathcal{G}(XY) + \mathcal{G}(XZ) - \mathcal{G}(XYZ) - \mathcal{G}(X) = 0$, with $Z = S - XY$.
   Next, for all x in $\pi_X(\varrho)$, let $\varrho^x = \pi_{S-X}(\varrho_x)$. We define the probability distribution p^x on ϱ^x by $p^x(t) = p(xt)/p_X(x)$ for all t in ϱ^x. Notice that p^x satisfies $p^x(t) \neq 0$ for all t in ϱ^x. Let $\mathcal{G}^x$ be the corresponding Gini index.

Now, let U be an arbitrary subset of S disjoint from X. We have that

$$\begin{aligned}
\mathcal{G}(XU) &= 1 - \sum_{xu \in \pi_{XU}(\varrho)} p^2_{XU}(xu) \\
&= 1 - \sum_{x \in \pi_X(\varrho)} \sum_{u \in \pi_U(\varrho_x)} p^2_X(x) {p^x_U}^2(x) \\
&= 1 - \sum_{x \in \pi_X(\varrho)} p^2_X(x) + \sum_{x \in \pi_X(\varrho)} p^2_X(x) \Big[1 - \sum_{u \in \pi_U(\varrho_x)} {p^x_U}^2(x) \Big] \\
&= \mathcal{G}(X) + \sum_{x \in \pi_X(\varrho)} p^2_X(x) \mathcal{G}^x(U).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathcal{G}(XY) + \mathcal{G}(XZ) - \mathcal{G}(XYZ) - \mathcal{G}(X) \\
&\qquad = \sum_{x \in \pi_X(\varrho)} p^2_X(x) \left[\mathcal{G}^x(Y) + \mathcal{G}^x(Z) - \mathcal{G}^x(S-X) - \mathcal{G}^x(\emptyset) \right].
\end{aligned}$$

Notice that, for all x in $\pi_X(\varrho)$, $\mathcal{G}^x(Y) + \mathcal{G}^x(Z) - \mathcal{G}^x(S-X) - \mathcal{G}^x(\emptyset) \geq 0$, since $\mathcal{G}^x$ is an isotone measure. Consequently, $\mathcal{G}(XY) + \mathcal{G}(XZ) - \mathcal{G}(XYZ) - \mathcal{G}(X) = 0$ if and only if, for all x in $\pi_X(\varrho)$, $\mathcal{G}^x(Y) + \mathcal{G}^x(Z) - \mathcal{G}^x(S-X) - \mathcal{G}^x(\emptyset) = \mathcal{G}^x(Y) + \mathcal{G}^x(Z) - \mathcal{G}^x(S-X) = 0$, or equivalently,

$$\sum_{y \in \pi_Y(\varrho^x)} {p^x_Y}^2(y) + \sum_{z \in \pi_Z(\varrho^x)} {p^x_Z}^2(z) - \sum_{t \in \varrho^x} p^{x2}(t) - 1 = 0.$$

If we rewrite the above equality using

$$\begin{aligned}
&\sum_{y \in \pi_Y(\varrho^x)} {p^x_Y}^2(y) = \sum_{y \in \pi_Y(\varrho^x)} \Big[\sum_{t \in \varrho^x \,\wedge\, t[Y]=y} p^x(t) \Big]^2, \\
&\sum_{z \in \pi_Z(\varrho^x)} {p^x_Z}^2(z) = \sum_{z \in \pi_Z(\varrho^x)} \Big[\sum_{t \in \varrho^x \,\wedge\, t[Z]=z} p^x(t) \Big]^2, \text{ and} \\
&1 = \Big[\sum_{t \in \varrho^x} p^x(t) \Big]^2,
\end{aligned}$$

we obtain, after simplification, that

$$\sum_{t,t' \in \varrho^x \,\wedge\, t[Y] \neq t'[Y] \,\wedge\, t[Z] \neq t'[Z]} p^x(t) p^x(t') = 0.$$

Notice, however, that $p^x(t) \neq 0$ for all t in ϱ^x. Hence, the above sum contains no terms. In other words, for all t, t' in ϱ^x, either $t[Y] = t'[Y]$ or $t[Z] = t[Z']$. This condition can only be satisfied if either $|\pi_Y(\varrho^x)| = 1$ or $|\pi_Z(\varrho^x)| = 1$, or, in other words, if ϱ_x satisfies either $X \overset{\text{fd}}{\to} Y$ or $X \overset{\text{fd}}{\to} Z$.
For the case that X and Y are *not* disjoint, notice that, from the rules in Proposition 14, it follows that $\mathcal{G}$ satisfies $X \twoheadrightarrow Y - X$. Hence, for all x in $\pi_X(\varrho)$, ϱ_x satisfies either $X \overset{\text{fd}}{\to} Y - X$ or $X \overset{\text{fd}}{\to} S - XY$. In the former case, ϱ_x also satisfies $X \overset{\text{fd}}{\to} Y$, by the augmentation rule for fds.
Finally, we observe that the condition that, for all x in $\pi_X(\varrho)$, $\sigma_{X=x}(\varrho)$ satisfies either $X \overset{\text{fd}}{\to} Y$ or $X \overset{\text{fd}}{\to} Z$ is *not* equivalent to ϱ satisfying either $X \overset{\text{fd}}{\to} Y$ or $X \overset{\text{fd}}{\to} Z$. Indeed, the relation ϱ given as

A	B	C
0	0	0
0	0	1
1	0	0
1	1	0

satisfies neither $A \overset{\text{fd}}{\to} B$ nor $A \overset{\text{fd}}{\to} C$, but $\sigma_{A=0}(\varrho)$ satisfies $A \overset{\text{fd}}{\to} B$ and $\sigma_{A=1}(\varrho)$ satisfies $A \overset{\text{fd}}{\to} C$.

Inexpressibility result We show that we cannot capture a multivalued dependency using any linear combination of the measure of its attributes using the Gini index measure. It is sufficient to show that any linear combination of the above cannot characterize the multivalued dependency $\emptyset \twoheadrightarrow A|B$. The equation we must satisfy is $a\mathcal{G}(A) + b\mathcal{G}(B) + c\mathcal{G}(AB) = 0$. We can assume for all practical purposes that $c = 1$. We show in Figure 1 three examples of the MVD $\emptyset \twoheadrightarrow A|B$. Now using the first two examples, we solve for the uknowns and then test our results on the third example. From the first example, we have $\mathcal{G}(AB) = \frac{3}{4}$, $\mathcal{G}(A) = \mathcal{G}(B) = \frac{1}{2}$, thus we get $\frac{1}{2}a + \frac{1}{2}b + \frac{3}{4} = 0$. From the second example, we have $\mathcal{G}(AB) = \mathcal{G}(A) = \frac{1}{2}$, $\mathcal{G}(B) = 0$, thus we get $\frac{1}{2}a + 0 + \frac{1}{2} = 0$. Solving these two equations we get that $a = -1, b = -\frac{1}{2}$. This means for an relation to satisfy the MVD $\emptyset \twoheadrightarrow A|B$ we must have $-\mathcal{G}(A) + -\frac{1}{2}\mathcal{G}(B) + \mathcal{G}(AB) = 0$. However, When applying this on the third example, with $\mathcal{G}(AB) = \mathcal{G}(B) = \frac{1}{2}$, $\mathcal{G}(A) = 0$, we get $\frac{-1}{4} + \frac{1}{2} \neq 0$. Thus a multivalued dependency cannot be expressed an any linear combination of a measure of its arributes when that measure is the Gini index.

A	B
1	1
1	2
2	1
2	2

A	B
1	1
2	1

A	B
1	1
1	2

Fig. 1. Examples of MVD $\emptyset \twoheadrightarrow A|B$

2. Let S be a relation scheme, let X and Y be subsets of S, and let ϱ be a nonempty relation instance over S. Let p be the *uniform* probability distribution over ϱ, satisfying $p(t) = 1/|\varrho|$ for all t in ϱ. The corresponding Shannon entropy measure $\mathcal{H}$ satisfies the secondary constraint $X \twoheadrightarrow Y$ if and only if ϱ satisfies the multivalued dependency $X \overset{\text{mvd}}{\twoheadrightarrow} Y$. (This fact was previously announced by Malvestuto [23], and rediscovered by Lee [30], and by Dalkilic and Robertson [12]. Related work can be found in [32]. We provide a proof of it here mainly to keep the paper self-contained.)
   To see this, we first assume that X and Y are disjoint.
   Observe that $\mathcal{H}$ satisfies $X \twoheadrightarrow Y$ if and only if $\mathcal{H}(XY)+\mathcal{H}(XZ)-\mathcal{H}(XYZ)-\mathcal{H}(X) = 0$, with $Z = S - XY$.
   Next, for all x in $\pi_X(\varrho)$, let $\varrho^x = \pi_{S-X}(\varrho_x)$. As before, we define the probability distribution p^x on ϱ^x by $p^x(t) = p(xt)/p_X(x)$ for all t in ϱ^x. Notice that $p^x(t) = 1/|\varrho^x|$ for all t in ϱ^x, whence p^x is the *uniform* probability distribution over ϱ^x. Let $\mathcal{H}^x$ be the corresponding Shannon measure.
   Now, let U be an arbitrary subset of S disjoint from X. We have that

$$\begin{aligned}
\mathcal{H}(XU) &= -\sum_{xu \in \pi_{XU}(\varrho)} p_{XU}(xu) \log p_{XU}(xu) \\
&= -\sum_{x \in \pi_X(\varrho)} \sum_{u \in \pi_U(\varrho^x)} p_X(x) p_U^x(u) \log[p_X(x) p_U^x(u)]
\end{aligned}$$

$$\begin{aligned}&= -\sum_{x\in\pi_X(\varrho)} p_X(x)\log p_X(x) \sum_{u\in\pi_U(\varrho^x)} p_U^x(u)\\&\qquad - \sum_{x\in\pi_X(\varrho)} p_X(x) \sum_{u\in\pi_U(\varrho^x)} p_U^x(u)\log p_U^x(u)\\&= \mathcal{H}(X) + \sum_{x\in\pi_X(\varrho)} p_X(x)\mathcal{H}^x(U)\end{aligned}$$

Hence,

$$\begin{aligned}&\mathcal{H}(XY)+\mathcal{H}(XZ)-\mathcal{H}(XYZ)-\mathcal{H}(X)\\&\qquad = \sum_{x\in\pi_X(\varrho)} p_X(x)\left[\mathcal{H}^x(Y)+\mathcal{H}^x(Z)-\mathcal{H}^x(S-X)-\mathcal{H}^x(\emptyset)\right].\end{aligned}$$

Notice that, for all x in $\pi_X(\varrho)$, $\mathcal{H}^x(Y)+\mathcal{H}^x(Z)-\mathcal{H}^x(S-X)-\mathcal{H}^x(\emptyset)\geq 0$, since $\mathcal{H}^x$ is an isotone measure. Consequently, $\mathcal{H}(XY)+\mathcal{H}(XZ)-\mathcal{H}(XYZ)-\mathcal{H}(X)=0$ if and only if, for all x in $\pi_X(\varrho)$, $\mathcal{H}^x(Y)+\mathcal{H}^x(Z)-\mathcal{H}^x(S-X)-\mathcal{H}^x(\emptyset) = \mathcal{H}^x(Y)+\mathcal{H}^x(Z)-\mathcal{H}^x(S-X)=0$. From the literature on Shannon's entropy measure [5, 11], it follows that the last equality holds if and only if, for all y in $\pi_Y(\varrho^x)$ and for all z in $\pi_Z(\varrho^x)$, $p^x(\sigma_{Y=y}\sigma_{Z=z}(\varrho^x)) = p^x(\sigma_{Y=y}(\varrho^x))p^x(\sigma_{Z=z}(\varrho^x))$. Using that p^x is the uniform probability distribution on ϱ^x, the last equality can be rewritten as

$$\frac{1}{|\varrho^x|} = \frac{|\sigma_{Y=y}(\varrho^x)|}{|\varrho^x|}\,\frac{|\sigma_{Z=z}(\varrho^x)|}{|\varrho^x|},$$

or $|\varrho^x| = |\sigma_{Y=y}(\varrho^x)|\,|\sigma_{Z=z}(\varrho^x)|$. Summing lefthand and righthand sides of the latter equality over all possible values of y and z yields

$$\sum_{y\in\pi_Y(\varrho^x)}\ \sum_{z\in\pi_Z(\varrho^x)} |\varrho^x| = \Big[\sum_{y\in\pi_Y(\varrho^x)} |\sigma_{Y=y}(\varrho^x)|\Big]\Big[\sum_{z\in\pi_Z(\varrho^x)} |\sigma_{Z=z}(\varrho^x)|\Big],$$

or $|\pi_Y(\varrho^x)|\,|\pi_Z(\varrho^x)|\,|\varrho^x| = |\varrho^x|^2$, or $|\pi_Y(\varrho^x)|\,|\pi_Z(\varrho^x)| = |\varrho^x|$. This condition expresses exactly that ϱ_x satisfies $X \overset{\text{mvd}}{\twoheadrightarrow} Y$. Furthermore, ϱ satisfies $X \overset{\text{mvd}}{\twoheadrightarrow} Y$ if and only if, for all x in $\pi_X(\varrho)$, ϱ_x satisfies $X \overset{\text{mvd}}{\twoheadrightarrow} Y$.

For the case that X and Y are *not* disjoint, notice that, from the rules in Proposition 14, it follows that $\mathcal{H}$ satisfies $X \twoheadrightarrow Y-X$.

Hence, ϱ satisfies $X \overset{\text{mvd}}{\twoheadrightarrow} Y - X$. By the augmentation rule for mvds, it follows that ϱ also satisfies $X \overset{\text{mvd}}{\twoheadrightarrow} Y$.

Inexpressibility result We show that we cannot capture a a cyclic join dependency using any linear combination of the measure of its attributes using the Entorpy measure. We show this by counter example. That is for a a given cyclic join dependency, it is not the case that $\sum_{u \subseteq \{A,B,C\}} a_i \mathcal{H}(u) = 0$. Consider the three cyclic join dependencies given in Figure 2. Observe that $\mathcal{H}(A) = \mathcal{H}(B) = \mathcal{H}(C)$ and $\mathcal{H}(AB) = \mathcal{H}(BC) = \mathcal{H}(AC)$ for all three relations (thus the coresponding constants for these measures are equal too). Thus we must satisfy the inequality $a\mathcal{H}(A) + b\mathcal{H}(AB) + \mathcal{H}(ABC) = 0$. For relation 1, we have $\mathcal{H}(A) = \frac{3}{4}\log\frac{4}{3} + 0.5$, $\mathcal{H}(AB) = \frac{3}{2}$, and $\mathcal{H}(ABC) = 2$. For relation 2, we have $\mathcal{H}(A) = 1$, $\mathcal{H}(AB) = 2$, and $\mathcal{H}(ABC) = 3$. Thus with this information we can solve for a and b to get the inequality, $4.08\mathcal{H}(A) - 3.54\mathcal{H}(AB) + \mathcal{H}(ABC) = 0$. When applying this inequality to the third cyclic join dependency (where $\mathcal{H}(A) = \frac{3}{5}\log\frac{5}{3} + \frac{2}{5}\log 5$, $\mathcal{H}(AB) = \frac{2}{5}\log\frac{5}{2} + \frac{3}{5}\log 5$, and $\mathcal{H}(ABC) = \log 5$) we get $4.08 * 1.37 - 3.54 * 1.92 + \log 5 \neq 0$. Thus a cyclic join dependency cannot be expressed an any linear combination of a measure of its arributes when that measure is the Shannon Entropy measure.

A	B	C
1	1	1
1	1	2
1	2	1
2	1	1

A	B	C
1	1	1
1	1	2
1	2	1
1	2	2
2	1	1
2	1	2
2	2	1
2	2	2

A	B	C
1	1	1
1	1	2
1	2	1
2	1	1
3	3	3

Fig. 2. Examples of cyclic join dependencies

Data mining—frequent item sets The frequency measure `freq` was described in Section 3.3, and in Section 6.1, it was shown that its associated primary constraints correspond to pure association rules. Its associated secondary constraints can be characterized as follows. For item sets X, Y, and Z, and basket set $\mathcal{B}$, $X \twoheadrightarrow Y|Z$ holds for $\texttt{freq}$ if and only if $\mathcal{B}(X) \subseteq \mathcal{B}(Y) \cup \mathcal{B}(Z)$. Indeed, $X \twoheadrightarrow Y|Z$ if and only if $\texttt{freq}''(X \twoheadrightarrow Y|Z) = 0$, which is the case if and only if $\texttt{freq}(XYZ) + \texttt{freq}(X) = \texttt{freq}(XY) + \texttt{freq}(XZ)$. This statement is equivalent to the following: $|\mathcal{B}(XYZ)| + |\mathcal{B}(X)| = |\mathcal{B}(XY)| + |\mathcal{B}(XZ)|$, or equivalently, $|\mathcal{B}(X) \cap \mathcal{B}(Y) \cap \mathcal{B}(Z)| + |\mathcal{B}(X)| = |\mathcal{B}(X) \cap \mathcal{B}(Y)| + |\mathcal{B}(X) \cap \mathcal{B}(Z)|$. The latter statement is equivalent with the condition $\mathcal{B}(X) \subseteq \mathcal{B}(Y) \cup \mathcal{B}(Z)$. We therefore call $X \twoheadrightarrow Y|Z$ a *disjunctive association rule*. Consequently, the strong secondary constraint $X \twoheadrightarrow Y$ holds if and only if $\mathcal{B}(X) \subseteq \mathcal{B}(Y) \cup \mathcal{B}(S - XY)$.

7 Completeness

In Section 6, we introduced measure constraints and specified inference rules. Their formal similarity to inference rules for functional and multivalued dependencies [4, 7] is apparent. Also in Section 6, we related satisfaction of measure constraints by some particular classes of measures to satisfaction of database constraints. We will exploit this to prove the completeness of some combinations of these rules for the inference of primary and strong secondary constraints.

Proposition 19. *The triviality, augmentation, and transitivity rules in Proposition 16 are sound and complete for the inference of primary constraints with respect to measures.*

Proof. Soundness follows from Proposition 12. To prove completeness, let $\mathcal{D}$ be a set of primary constraints, let $X \to Y$ be a primary constraint, and assume that each measure satisfying all constraints in $\mathcal{D}$ also satisfies $X \to Y$. Now, let ϱ be an arbitrary relation satisfying all functional dependencies in $\mathcal{D}_{\mathrm{fd}} = \{V \stackrel{\mathrm{fd}}{\to} Y \mid V \to Y \in \mathcal{D}\}$. If ϱ is empty, ϱ trivially satisfies $X \stackrel{\mathrm{fd}}{\to} Y$. If ϱ is not empty, we proceed as follows. Let p be the uniform probability distribution on ϱ. By the observations above, the Shannon measure $\mathcal{H}$ corresponding to p satisfies all constraints in $\mathcal{D}$. By assumption, $\mathcal{H}$ also satisfies $X \to Y$. Hence, ϱ satisfies $X \stackrel{\mathrm{fd}}{\to} Y$. We have thus proved that the functional dependency $X \stackrel{\text{fd}}{\to} Y$ is logically implied by the set of functional dependencies $\mathcal{D}_{\text{fd}}$. As shown by Armstrong [4], $X \stackrel{\text{fd}}{\to} Y$ can be inferred from $\mathcal{D}_{\text{fd}}$ by the reflexivity, augmentation and transitivity rules for functional dependencies. Hence, $X \to Y$ can be inferred from $\mathcal{D}$ by the triviality, augmentation, and transitivity rules for primary constraints.

Proposition 20. *The triviality, augmentation, and transitivity rules for primary constraints in Proposition 16, together with the triviality, complementation, augmentation, and pseudotransitivity rule for strong secondary constraints in Proposition 17 are sound and complete for the inference of primary and strong secondary constraints together with respect to measures.*

The proof is completely analogous to the proof of Proposition 19, except that we do no longer have the choice between using the Gini index or the Shannon measure: we must use the latter, now.

8 Future directions

A possible direction for future work is to consider other measures that fit in our setting. One class of such measures is that based on Tsallis' entropy [31]. These measures generalize both the Gini index and the Shannon entropy measure and were studied in the context of databases by Simovici and Jaroszewics [28]. In that work, the concept of conditional entropy arose and we plan to relate our work to it.

In many cases, the definition of a measure occurs in the context of another structure, for example, a relation in the definition of the Gini index and the Shannon measure, and a collection of baskets in the definition of the frequency measure. In such cases, one can think of a measure as a two-parameter function $\mathcal{M}(X, s)$ wherein X is a subset of S and s is some structure over S. In this view, it is natural to study such measures letting the parameter s vary. Hilderman and Hamilton[19] conducted such as study for measures including the Simpson and Shannon measures, for the case wherein X is chosen to be S. It would be useful to consider situations wherein both X and s vary simultaneously.

As is well known, the theory of relational design is centered around functional and multivalued dependencies. Concepts such as keys, closures, normal forms, etc. can be defined in terms of them [22]. Since primary and secondary constraints are generalizations of these constraints, it is natural to re-interpret these concepts in the more general setting.[10]

Finally, this paper considers measures that are defined in terms of properties of their first and second finite differentials. An obvious idea is to study measures in terms of properties of their higher-order finite differentials. Initial work in that direction, related to inference systems, was reported in [26].

Acknowledgments: We thank Paul Purdom and Edward Robertson for helpful discussions on topics covered in this paper.

References

1. R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proc. 20th Int. Conf. Very Large Data Bases, VLDB*, pages 487–499, 1994.
2. P. Andritsos, R. Miller, and P. Tsaparas. Information-theoretic tools for mining database structure from large data sets. In *Proc.SIGMOD*. ACM Press, 2004.
3. M. Arenas and L. Libkin. An information-theoretic approach to normal forms for relational and xml data. In *Proc. SIGMOD*, pages 15–26. ACM Press, 2003.
4. W. Armstrong. Dependency structures of database relationships. In *In Proceedings of the tri-annual IFIP Congress*, pages 580–583, 1974.
5. R.A. Ash. *Information Theory.* Dover Publications, Inc. New York, 1965.
6. R. Baeza-Yates and B. Ribeirto-Neto. *Modern Information Retrieval.* Addison-Wesley-Longman, 1999.
7. C. Beeri, R. Fagin, and J. Howard. A complete axiomatization for functional and multivalued dependencies in database relations. In *Proc. SIGMOD*, pages 47–61, 1977.
8. S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30:107–117, 1998.
9. T. Calders. *Axiomatization and Deduction Rules for the Frequency of Itemsets.* PhD dissertation- University of Antwerp, 2003.
10. D. Cohn. *Measure Theory.* Birkhäuser-Boston, 1980.
11. T.M. Cover and J.A. Thomas. *Elements of Information Theory.* Wiley-Interscience Publication, 1991.
12. M. Dalkilic and E. Robertson. Information dependencies. In *Symposium on Principles of Database Systems*, pages 245–253, 2000.

[10] This may lead to design principles in new contexts. Examples of such design principles are based on the idea of defining good data designs in terms of maximizing information content. Such ideas have been explored in [3] and [2].

13. V. Diodato. *Dictionary of Bibliometrics.* Haworth Library and Information Science, 1994.
14. F. Fujishige. Polymatroidal dependence structure of a set of random variables. *Information and Control*, 39:55–72, 1978.
15. S. Fujishige. Submodular functions and optimization. *Annals of Discrete Mathematics*, 47, 1991.
16. C. Gini. Measurement of inequality of incomes. *The Economic Journal*, 31:124–126, 1921.
17. S. Goldberg. *Introduction to Difference Equations.* Dover, 1986.
18. J. Gray, S. Chaudhuri, A. Bosworth, A. Layman, D. Reichart, M. Venkatrao, F. Pellow, and H. Pirahesh. Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. *J. Data Mining and Knowledge Discovery*, 1(1):29–53, 1997.
19. R. Hilderman and H. Hamilton. Principles for mining summaries using objective measures of interestingness. In *Proceedings of the Twelfth IEEE International Conference on Tools with Artificial Intelligence (ICTAI'00)*, pages 72–81. Vancouver, Canada, November, 2000.
20. V. Krishnamurthy. *Combinatorics: Theory and Applications.* Ellis Horwood Limited, 1986.
21. E. H. Lieb and M. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14(12):1938–1941, 1973.
22. D. Maier. *The Theory of Relational Databases.* Computer Science Press, 1983.
23. F. M. Malvestuto. Statistical treatment of the information content of a database. *Information Systems*, 11:211–223, 1986.
24. P.E. Pfeiffer. *Concepts of Probability Theory (2nd revised edition).* Dover Publications, Inc., 1979.
25. V. Poosala, P. Haas, Y. Ioannidis, and E. Shekita. Improved histograms for selectivity estimation of range predicates. In *Proceedings of the 1996 International Conference on Management of Data (SIGMOD)*, pages 294–305, 1996.
26. B. Sayrafi and D. Van Gucht. Inference systems derived from additive measures. *Workshop on Causality and Causal Discovery, London, Canada*, 2004.
27. C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October 1948.
28. D.A. Simovici and S. Jaroszewics. Generalized entropy and decision trees. *EGC 2003 - Journees francophones d'Extraction et de Gestion de Connaissances*, pages 369–380, 2003.
29. E. Simpson. Measurement of diversity. In *Nature*, volume 163, page 688, 1949.
30. T. Lee. An Information-Theoretic Analysis of Relational Databases – Part I: Data Dependencies and Information Metric. *IEEE Transactions on Software Engineering*, SE-13:1049–1061, 1987.
31. P. Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics*, 52:479–87, 1988.
32. S.K.M. Wong, C.J. Butz, and D. Wu. On the implication problem for probabilistic conditional independency. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 30.6:785–805, 2000.