TECHNICAL REPORT NO. 531

Information Dependencies

by
Mehmet M. Dalkilic
Edward L. Robertson

November 1999

COMPUTER SCIENCE DEPARTMENT
INDIANA UNIVERSITY
Bloomington, Indiana 47405-4101

Information Dependencies

Mehmet M. Dalkilic
Indiana University Computer Science
Lindley Hall 215
Bloomington, IN 47405 USA
812-855-4318
dalkilic@cs.indiana.edu

Edward L. Robertson
Indiana University Computer Science
Lindley Hall 215
Bloomington, IN 47405 USA
812-855-4954
edrbtsn@cs.indiana.edu

Abstract

This paper uses the tools of information theory to examine and reason about the information content of the attributes within a relation instance. For two sets of attributes, an *information dependency measure* (InD measure) characterizes the uncertainty remaining about the values for the second set when the values for the first set are known. A variety of arithmetic inequalities (InD *inequalities*) are shown to hold among InD measures; InD inequalities hold in any relation instance. Numeric constraints (InD *constraints*) on InD measures, consistent with the InD inequalities, can be applied to relation instances. Remarkably, functional and multivalued dependencies correspond to setting certain constraints to zero, with Armstrong's axioms shown to be consequences of the arithmetic inequalities applied to constraints. As an analog of completeness, for any set of constraints consistent with the inequalities, we may construct a relation instance that approximates these constraints within any positive ϵ.

1 Introduction

That the well-developed discipline of *information theory* seemed to have so little to say about *information systems* is a long-standing conundrum. Attempts to use information theory to "measure" the information content of a relation are blocked by the inability to accurately characterize the underlying domain. An answer to this mystery is that we have been looking in the wrong place. The tools of information theory, dealing closely with representation issues, apply within a relation instance and between the various attributes of that instance.

The traditional approach to information theory is based upon communication via a *channel*. In each instance there is a fixed set of messages $M = \{v_1, \ldots, v_n\}$; when one of these is transmitted from the *sender* to the *receiver* (via the channel), the receiver gains a certain amount of information. The less likely a message is to be sent, the more meaningful is its receipt. This is formalized by assigning to each message v_i a probability p_i (subject to the natural constraint that $\sum_{i=1}^{n} p_i = 1$) and defining the information content of v_i to be $\log {}^1/_{p_i}$ (all logarithms in this paper are base 2).

Another way of viewing this measure is that the amount of information in a message is related to how "surprising" the message is–a weather report during the month of July contains little information if the prediction is "hot," but a prediction of "snow" carries a lot of information. The issue of surprise is also related to the recipient's "state of knowledge." In the weather report example, the astonishment of the report "snow" was directly related to the knowledge that it was July; in January the information content of the two reports would be vastly different. Thus

the in– or inter–dependence of two sets of messages is highly significant. If two message sets are independent (in the intuitive and the statistical sense), receipt of a message from one set does not alter the information content of the other (*e.g.* temperature and wind speed). If two message sets are not independent, receipt of a message from the first set may greatly alter the likelihood of receipt, and hence information content, of messages from the second set (*e.g.* temperature and form of precipitation).

A central concept in information theory is the *entropy* $\mathcal{H}$ of a set of messages, the weighted average of the message information:

Definition 1.1 *Entropy.* Given a set $M = \{v_1, \ldots, v_n\}$ of messages with probabilities $P_M = \{p_1, \ldots, p_n\}$, the entropy of M is $\mathcal{H}(M) = \mathcal{H}(p_1, \ldots, p_n) = \sum_{i=1}^{n} p_i \log {}^1/_{p_i}$

Entropy is closely related to encoding of messages, in that encoding each v_i using $\log {}^1/_{p_i}$ bits gives the minimal number of expected bits for transmitting messages of M.

Remark 1 Suppose for messages of M, no probability is 0 and $\mathcal{H}(M) = 0$. Then $M = \{v_1\}$, *i.e.* M contains a single message.

In a database context, information content is measured in terms of selection (specification of a specific value) rather than transmission. This avoids the thorny problem which seems to say that, since the database is stored on site and no transmission occurs, there is no information. In particular, the model looks at an instance of a single relation and at values for some arbitrarily selected tuple. For simplicity, we assume that the message source is ergodic–all tuples are equally likely; a probability distribution could be applied to the tuples with less impact on the formalism than on the intuition. Because of the assumption that all tuples are equally likely, the information required to specify one particular tuple from a relation instance with n tuples is, of course, $\log n$ and the minimal cost of encoding requires uniformly $\log n$ bits. We treat an attribute A as the equivalent of a message source, where the message set is the active domain and each value v_i has probability ${c_i}/{n}$, when v_i occurs c_i times. Thus a single value carries $\log n$ bits only if it is drawn from an attribute which has n distinct values, that is, when the attribute is a key. The class standing code at a typical four–year college has approximately two bits of information (somewhat less, to the extent that attrition has skewed enrollment) while gender at VMI has little information (using the entropy measure, since the information content of the value "female" is high, but its receipt is unlikely).

The major results of this paper use the common definition of information to characterize *information dependencies*. This characterization has three steps. The first extends the use of entropy as a measure of information to be an *information dependency measure* (Section 4). The second derives a number of *arithmetic inequalities* which must always hold between particular measures in a relation instance (Section 5). The third investigates the consequences of placing *numeric constraints* on some or all measures of a relation instance. Most significantly, functional and multi-valued dependency result from constraining certain particular measures (or their differences) to zero (Section 6).

For example, in a weather report database, `month` has entropy 3.58 and we might discover that `condition` has entropy 1.9. But in a fixed month, `condition` has entropy approximately 1.6. Thus knowing the value of `month` contributes approximately 0.3 bits of information to knowledge of `condition`, with 1.6 bits of uncertainty. On the other hand, in a personnel database where `EmpID` $\rightarrow$ `salary`, `EmpID` provides the entire information content of `salary` with 0 bits uncertain.

In addition, the measure/constraint formulation exhibits an analog of completeness in that, for any set of numeric constraints consistent with the arithmetic inequalities and any positive ϵ, there is a relation instance that achieves those constraints within ϵ (Section 7).

This characterization of information dependency has many important theoretic and practical implications. It allows us to more carefully investigate notions of approximate functional dependency. It can help with normalization. It opens up whole realms of data mining approaches.

2 Preliminaries

Here are the notations and conventions.

Relations All relation instances are non-empty and multi-sets. $\mathbf{r}$, $\mathbf{s}$ denote instances[1]. Operators π, σ do not filter for distinctiveness.

Attributes R is schema for instance $\mathbf{r}$ and $X, Y, Z, V, W \subseteq R$. XY denotes $X \cup Y$ and A is equivalent to $\{A\}$ for $A \in R$. X, Y, Z partition R.

Values v is equivalent to $\langle v \rangle$ when $\langle v \rangle \in \pi_A(\mathbf{r})$. $\ell = \mathbf{count\text{-}distinct}(\pi_X(\mathbf{r}))$. x_i enumerates the instances of $\mathbf{distinct}(\pi_X(\mathbf{r}))$, so $1 \leq i \leq \ell$, similarly for m and y_j *wrt* Y, and n and z_k *wrt* Z.

Probabilities $P(S = v) = \frac{\mathbf{count}(\sigma_{S=v}(\mathbf{r}))}{\mathbf{count}(\mathbf{r})}$ for $S \subseteq R$. $p_i = P(X = x_i)$ (note use of i is consistent with above), similarly $p_j = P(Y = y_j)$, $p_k = P(Z = z_k)$, $p_{ij} = P(X = x_i \& Y = y_j)$, and so forth. $\sum_i^n p_i = \sum_{i=1}^n p_i$ and likewise for p_j, p_k, *etc.*

Two central notions to entropy are conditional probability and statistical independence. Conditional probability allows us to make a possibly more informed probability measure of a set of values by narrowing the scope of overall possibilities. Independence establishes a bound on how informed the conditional probability enables us to be.

Definition 2.1 *Conditional Probability.* The *conditional probability* of $Y = y_j$ given $X = x_i$, written $P(Y = y_j | X = x_i)$, is $P(Y = y_j)$ in the instance $\sigma_{X=x_i}(\mathbf{r})$. In symbols,
$P(Y = y_j | X = x_i) = P(X = x_i \,\&\, Y = y_j)/P(X = x_i)$.

Definition 2.2 *Independence.* X, Y are *independent* if $P(Y = y_j) = P(Y = y_j | X = x_i)$.

In this paper, there are log function expressions of the form $\log(1/0)$. By convention (continuity arguments), $0 \log(1/0) = 0$. and $a \log(1/0) = +\infty$ for real number $a > 0$.

Lemma 2.1 $\log x \leq x - 1$.

Lemma 2.2 Let $P = \{p_1, \ldots, p_n\}$ be a probability distribution and $Q = \{q_1, \ldots, q_n\}$ such that $\sum_i^n q_i \leq 1$ and $(\forall\, i)\ 1 \leq i \leq n,\ 0 \leq q_i \leq 1$. Then $\sum_i^n p_i \log {}^1\!/_{p_i} \leq \sum_i^n p_i \log {}^1\!/_{q_i}$.

PROOF

$$
\begin{array}{rcll}
\log {}^{q_i}\!/_{p_i} & \leq & {}^{q_i}\!/_{p_i} - 1 & \textbf{Lm 2.1} \\
p_i \log {}^{q_i}\!/_{p_i} & \leq & q_i - p_i & \\
p_i \log {}^1\!/_{p_i} & \leq & p_i \log {}^1\!/_{q_i} + q_i - p_i & \\
\sum_i^n p_i \log {}^1\!/_{p_i} & \leq & \sum_i^n (p_i \log {}^1\!/_{q_i} + q_i - p_i) & \\
\sum_i^n p_i \log {}^1\!/_{p_i} & \leq & \sum_i^n p_i \log {}^1\!/_{q_i} + \sum_i^n q_i - \sum_i^n p_i \quad = \quad \sum_i^n p_i \log {}^1\!/_{q_i} + \delta - 1 & \text{where } \delta \leq 1 \\
 & \leq & \sum_i^n p_i \log {}^1\!/_{q_i} &
\end{array}
$$

[1] *Null* values are not considered here.

3 The bounds on entropy

To ease notation, we write $\mathcal{H}_X$ for $\mathcal{H}(X)$. From now on, we understand that $\mathcal{H}$ is always associated with a non-empty instance $\mathbf{r}$. When $\mathbf{r}$ is not clear from context, we write $\mathcal{H}_X^{\mathbf{r}}$. In the remainder of this section, we establish upper and lower bounds on the entropy function.

Lemma 3.1 *Upper and Lower Bounds on Entropy.* $0 \leq \mathcal{H}_X \leq \log \ell$.

PROOF Since $0 \leq p_i \leq 1$, $\log p_i \leq 0 \Rightarrow \log {}^1\!/_{p_i} \geq 0$; consequently, $\mathcal{H}_X \geq 0$. Suppose $p_i = {}^1\!/_{\ell}$. By Lemma 2.2, $\log {}^1\!/_{\ell p_i} \leq {}^1\!/_{\ell p_i} - 1 \Rightarrow \sum_i^{\ell} p_i \log {}^1\!/_{p_i} \leq \sum_i^{\ell} (p_i \log k + {}^1\!/_k - p_i) \leq \log k$.

Intuitively, the entropy of a set $X \subseteq R$ equal to zero signifies that there exists no uncertainty or information, whereas, equal to $\log \ell$ signifies complete uncertainty or information. A consequence of our notation allows us to find the joint entropy of sets $X, Y \subseteq R$. The joint entropy of X, Y, written $\mathcal{H}_{XY}$, is $\mathcal{H}_{XY} = \mathcal{H}(p_{1,1}, \ldots, p_{\ell,m}) = \sum_i^{\ell} \sum_j^{m} p_{i,j} \log {}^1\!/_{p_{i,j}}$.

Lemma 3.2 *Bounds on Joint Entropy.* $X, Y \subseteq R$,

$$\mathcal{H}_X + \mathcal{H}_Y \geq \mathcal{H}_{XY} \geq \mathbf{max}(\mathcal{H}_X, \mathcal{H}_Y)$$

with $\mathcal{H}_X + \mathcal{H}_Y = \mathcal{H}_{XY}$ if X, Y are independent.

PROOF First inequality:

$$\begin{array}{rcll} \mathcal{H}_X + \mathcal{H}_Y & = & \sum_i^{\ell} p_i \log {}^1\!/_{p_i} + \sum_j^{m} \log {}^1\!/_{p_j} & = \sum_i^{\ell} \sum_j^{m} p_{i,j} \log {}^1\!/_{(p_i \cdot p_j)} \\ & \geq & \sum_i^{\ell} \sum_j^{m} p_{i,j} \log {}^1\!/_{p_{i,j}} & \textbf{Lm 2.2} \\ & = & \mathcal{H}_{XY} & \end{array}$$

When X and Y are independent, $p_{i,j} = p_i \cdot p_j$ and thus, the inequality in the above deduction is in fact equality.

Second inequality: Observe that $p_i = \sum_j p_{i,j}$. Let $q_i = \mathbf{max}\{p_{i,j} | 1 \leq j \leq m\}$. Then for any j, $p_i \geq q_i \geq p_{i,j}$ and consequently, $\log {}^1\!/_{p_{i,j}} \geq \log {}^1\!/_{q_i}$ and thus,

$$\begin{array}{l} \mathcal{H}_X \\ \quad = \sum_i^{\ell} p_i \log {}^1\!/_{p_i} \leq \sum_i^{\ell} p_i \log {}^1\!/_{q_i} \ \textbf{Lm 2.2} \\ \quad = \sum_i^{\ell} \sum_j^{m} p_{i,j} \log {}^1\!/_{q_i} \leq \sum_i^{\ell} \sum_j^{m} p_{i,j} \log {}^1\!/_{p_{i,j}} \\ \quad = \mathcal{H}_{XY} \end{array}$$

and symmetrically for $\mathcal{H}_Y$ as well.

4 InD measures

An *information dependency measure* (InD measure) between X and Y, for $X, Y \subseteq R$, attempts to answer the question "How much do we *not* know about Y provided we know X?" Using the notation of **Section** 2, if we know that $X = x_i$, then we are possibly more informed about $Y = y_j$ and therefore, can recalculate the entropy of Y as

$$\begin{array}{l} \mathcal{H}(Y | X = x_i) = \\ \quad \mathcal{H}(\, {}^{p_{i,1}}\!/_{p_i}, \ldots, {}^{p_{i,m}}\!/_{p_i}\,) = \sum_j^{m} \frac{p_{i,j}}{p_i} \log \frac{p_i}{p_{i,j}} \end{array}$$

Amortizing this over each of the ℓ different X values according to the respective probabilities p_i gives the entropy of Y dependent on X, resulting in the following definition of an information dependency measure. Note that these are measures, not metrics.

r

A	B
a	e
a	f
a	e
a	f
b	g
b	g
c	g
d	g

InD measures

$$\begin{aligned} \mathcal{H}_{\mathtt{A}} &= 7/4 \\ \mathcal{H}_{\mathtt{B}} &= 3/2 \\ \mathcal{H}_{\mathtt{AB}} &= 9/4 \\ \mathcal{H}_{\mathtt{A}\rightarrow\mathtt{B}} &= 1/2 \\ \mathcal{H}_{\mathtt{B}\rightarrow\mathtt{A}} &= 3/4 \end{aligned}$$

Figure 1: (**left**) An instance **r**. (**right**) InD measures of **r**. Observe that $\mathcal{H}_{\mathtt{A}\rightarrow\mathtt{B}} = \mathcal{H}_{\mathtt{AB}} - \mathcal{H}_{\mathtt{A}}$ and $\mathcal{H}_{\mathtt{A}\rightarrow\mathtt{B}} = \sum_i^4 p_{a_i} \cdot \mathcal{H}(\mathtt{B}|a_i) = 1/2\,\mathcal{H}(1/2,1/2,0) + 1/4\,\mathcal{H}(0,0,1) + 2(1/8)\,\mathcal{H}(0,0,1) = 1/2 + 0 + 0 = 1/2$.

Definition 4.1 *Information Dependency Measure.* The *information dependency measure* (InD measure) of Y given $X$is $\mathcal{H}_{X\rightarrow Y}$,

$$\textstyle\sum_i^\ell p_i \cdot \mathcal{H}(Y|X = x_i) = \sum_i^\ell \sum_j^m p_{i,j} \log \frac{p_i}{p_{i,j}}$$

We will now normally drop the word "entropy" when referring to these measures, but that this value is not a declaration of dependency (as is the case with FDs) but a measure of dependency is important to keep in mind. We now characterize an InD measure $\mathcal{H}_{X\rightarrow Y}$ in terms of InD measures $\mathcal{H}_X$ and $\mathcal{H}_{XY}$.

Lemma 4.1 $\mathcal{H}_{X\rightarrow Y} = \mathcal{H}_{XY} - \mathcal{H}_X$.

PROOF

$$\begin{aligned} &\mathcal{H}_{X\rightarrow Y} \\ &= \textstyle\sum_i^\ell \sum_j^m p_{i,j} \log p_i/p_{i,j} \\ &\quad = \textstyle\sum_i^\ell \sum_j^m p_{i,j}[\log 1/p_{i,j} - \log 1/p_i\,] \\ &= \textstyle\sum_i^\ell \sum_j^m p_{i,j} \log 1/p_{i,j} - \sum_i^\ell \sum_j^m p_{i,j} \log 1/p_i \\ &= \mathcal{H}_{XY} - \textstyle\sum_i^\ell p_i \log 1/p_i \\ &= \mathcal{H}_{XY} - \mathcal{H}_X \end{aligned}$$

Note that $\mathcal{H}_{X\rightarrow Y}$ is a measure of the information needed to represent Y given that X is known, not the information that X contains about Y. This latter quantity of course is measured by $\mathcal{H}_Y - \mathcal{H}_{X\rightarrow Y}$.

5 InD measure inequalities

The relationships among InD measures are characterized by inequalities and expressions involving the various measures. Of these formulae, several are named according to the corresponding functional dependency inference rules, which they characterize under special circumstances.

Lemma 5.1 *Reflexivity.* $H_{X\rightarrow Y} = 0$, *for* $Y \subseteq X \subseteq R$.
PROOF Let $Z = YX$. Then by **Lm** 4.1 $\mathcal{H}_{Z\rightarrow Y} = \mathcal{H}_{ZY} - \mathcal{H}_Z = \mathcal{H}_Z - \mathcal{H}_Z = 0$.

Lemma 5.2 $H_{XZ\rightarrow YZ} = H_{XZ\rightarrow Y}$.

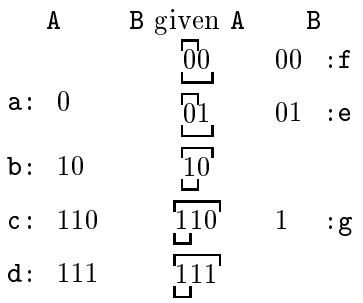

Figure 2: Encodings of `A`, `B`, `B` given `A` from Fig.1. The $\sqcap$ contains the portion of the bit string that encodes `A`, $\sqcup$ similarly for `B`. Where $\begin{smallmatrix}\sqcap\\ \sqcup\end{smallmatrix}$ overlap shows the portion of the encoding of `B` that is contained within the encoding of `A`. The surprise after receiving `A=a` is witnessed by the fact that, although we know we will receive the first bit of `B=e` or `B=f`, *i.e.* 0, we need an additional 1/4 bits for both the second bit of `B=e` and `B=f`. Receipt of `A=b`,`A=c`, or `A=d`, on the other hand, poses no surprise since `B=g` is completely contained therein.

PROOF

$$\begin{aligned}
&\mathcal{H}_{XZ\to YZ}\\
&= \mathcal{H}_{XYZ} - \mathcal{H}_{XZ} && \textbf{Lm } 4.1\\
&= \mathcal{H}_{XZ\to Y} + \mathcal{H}_{XZ} - \mathcal{H}_{XZ} && \textbf{Lm } 4.1\\
&= \mathcal{H}_{XZ\to Y}
\end{aligned}$$

illustrate the situation: two InDs may interact little so they combine to sum their InDs, or they may interact strongly, so their combination yields total dependencies. Putting restrictions on the left- or right-hand sides constrains the interactions and hence tightens the InD relationships.

Lemma 5.3 *Union* (left). $\mathcal{H}_{X\to Y} + \mathcal{H}_{X\to Z} \geq \mathcal{H}_{X\to YZ}$ *with equality if* $p_{j|i}$ *and* $p_{k|i}$ *are independent.*

PROOF

$$\begin{aligned}
&\mathcal{H}_{X\to Y} + \mathcal{H}_{X\to Z}\\
&= \textstyle\sum_i^n \sum_j^m p_{i,j} \log {}^1\!/_{p_{j|i}} + \sum_i^n \sum_k^q p_{i,k} \log {}^1\!/_{p_{k|i}}\\
&= \textstyle\sum_i^n \sum_j^m \sum_k^q p_{i,j,k} [\log {}^1\!/_{p_{j|i}} + \log {}^1\!/_{p_{k|i}}]\\
&= \textstyle\sum_i^n p_i \sum_j^m \sum_k^q p_{j,k|i} \log {}^1\!/_{p_{j|i}\, p_{k|i}}\\
&\geq \textstyle\sum_i^n p_i \sum_j^m \sum_k^q p_{j,k|i} \log {}^1\!/_{p_{j,k|i}}\\
&\quad (\forall i)\, 1 \leq i \leq n, \ \textbf{Lm } 2.2
\end{aligned}$$

Lemma 5.4 $\mathcal{H}_{X\to YZ} = \mathcal{H}_{X\to Y} + \mathcal{H}_{XY\to Z} \geq \mathbf{max}(\mathcal{H}_{X\to Y}, \mathcal{H}_{XY\to Z})$.

PROOF

$$\begin{aligned}
&\mathcal{H}_{X\to YZ}\\
&= \mathcal{H}_{XYZ} - \mathcal{H}_{X} && \textbf{Lm } 4.1\\
&= \mathcal{H}_{XY\to Z} + \mathcal{H}_{XY} - \mathcal{H}_{X} && \textbf{Lm } 4.1\\
&= \mathcal{H}_{XY\to Z} + \mathcal{H}_{X\to Y} && \textbf{Lm } 4.1
\end{aligned}$$

Lemma 5.5 $\mathcal{H}_{XY\to Z} \leq \mathcal{H}_{X\to Z}$.

PROOF

$$\begin{aligned}
&\mathcal{H}_{XY\to Z}\\
&= \mathcal{H}_{X\to YZ} - \mathcal{H}_{X\to Y} && \textbf{Lm } 5.4\\
&\leq \mathcal{H}_{X\to Y} + \mathcal{H}_{X\to Z} - \mathcal{H}_{X\to Y} && \textbf{Lm } 5.3\\
&\leq \mathcal{H}_{X\to Z}
\end{aligned}$$

Lemma 5.6 *Union* (right). $\mathbf{min}(\mathcal{H}_{X\to Z}, \mathcal{H}_{Y\to Z}) \geq \mathcal{H}_{XY\to Z}$.
PROOF

$\mathcal{H}_{X\to Z} \geq \mathcal{H}_{XY\to Y}$ **Lm** 5.5
$\mathcal{H}_{Y\to Z} \geq \mathcal{H}_{XY\to Z}$ **Lm** 5.5
$\mathbf{min}(\mathcal{H}_{X\to Z}, \mathcal{H}_{Y\to Z}) \geq \mathcal{H}_{XY\to Z}$

Lemma 5.7 *Augmentation* (1). $\mathcal{H}_{XZ\to YZ} \leq \mathcal{H}_{X\to Y}$.
PROOF

$\mathcal{H}_{XZ\to YZ}$
$= \mathcal{H}_{XZ\to Y}$ **Lm** 5.2
$\leq H_{X\to Y}$ **Lm** 5.5

Lemma 5.8 *Transitivity.* $H_{X\to Y} + H_{Y\to Z} \geq H_{X\to Z}$.
PROOF

$\mathcal{H}_{X\to Y} + \mathcal{H}_{Y\to Z}$
$\geq \mathcal{H}_{X\to XY} + \mathcal{H}_{XY\to XZ}$ **Lm** 5.7
$= \mathcal{H}_{XY} - \mathcal{H}_X + \mathcal{H}_{XYZ} - \mathcal{H}_{XY}$ **Lm** 4.1
$= \mathcal{H}_{XYZ} - \mathcal{H}_X$
$\geq \mathcal{H}_{XZ} - \mathcal{H}_X$ **Lm** 3.2
$= \mathcal{H}_{X\to Z}$ **Lm** 4.1

Lemma 5.9 *Union* (full). $\mathcal{H}_{X\to Y} + \mathcal{H}_{W\to Z} \geq \mathcal{H}_{XW\to YZ}$
PROOF

$\mathcal{H}_{X\to Y} + \mathcal{H}_{W\to Z}$
$\geq \mathcal{H}_{XW\to YW} + \mathcal{H}_{WY\to ZY}$ **Lm** 5.7
$\geq \mathcal{H}_{XW\to YZ}$ **Lm** 5.8

Lemma 5.10 *Decomposition.* if $Z \subseteq Y$, then $\mathcal{H}_{X\to Y} \geq \mathcal{H}_{X\to Z}$.
PROOF

$\mathcal{H}_{Y\to Z} = 0$ **Lm** 5.1
$\mathcal{H}_{X\to Y} + \mathcal{H}_{Y\to Z} \geq \mathcal{H}_{X\to Z}$ **Lm** 5.8
$\mathcal{H}_{X\to Y} \geq \mathcal{H}_{X\to Z}$

Lemma 5.11 *Psuedotransitivity.* $\mathcal{H}_{X\to Y} + \mathcal{H}_{WY\to Z} \geq \mathcal{H}_{XW\to Z}$.
PROOF

$\mathcal{H}_{X\to Y} + \mathcal{H}_{WY\to Z}$
$\geq \mathcal{H}_{XW\to YW} + \mathcal{H}_{WY\to Z}$ **Lm** 5.7
$\geq \mathcal{H}_{XW\to Z}$ **Lm** 5.8

Lemma 5.12 For $XYZ = R$, if $\mathcal{H}_{X\to Y} + \mathcal{H}_{X\to Z} = \mathcal{H}_{X\to YZ}$, then $\mathcal{H}_{WX\to YV} + \mathcal{H}_{WX\to Z-V} = \mathcal{H}_{WX\to YZ}$.
PROOF By **Lm** 5.2 we may assume *wlog* $V \subseteq W \subseteq Y \cup Z$. Let $\check{Y} = W \cap Y$ and $\check{Z} = W \cap Z$.

$\mathcal{H}_{X\to Y} + \mathcal{H}_{X\to Z} - \mathcal{H}_{X\to YZ}$
$= \mathcal{H}_{XY} - \mathcal{H}_X + \mathcal{H}_{XZ} - \mathcal{H}_X - \mathcal{H}_{XYZ} + \mathcal{H}_X$
$\geq \mathcal{H}_{XY\check{Y}} - \mathcal{H}_{X\check{Z}} + \mathcal{H}_{XZ\check{Z}} - \mathcal{H}_{XYZ}$
$\geq \mathcal{H}_{XY\check{Y}\check{Z}} + \mathcal{H}_{XZ\check{Y}\check{Z}} - \mathcal{H}_{X\check{Y}\check{Z}} - \mathcal{H}_{XYZ}$
$= \mathcal{H}_{XYWV} + \mathcal{H}_{X(Z-V)W} - \mathcal{H}_{XW} - \mathcal{H}_{WYZW}$
$= \mathcal{H}_{WX\to YV} + \mathcal{H}_{WX\to(Z-V)} - \mathcal{H}_{WX\to YZ}$

6 FDs, MVDs, and Armstrong's axioms

6.1 Functional dependencies

Functional dependencies (FDs) are long–known and well–studied [8, 10]. For $X, Y \subseteq R$, X *functionally determines* Y, written $X \rightarrow Y$, if any X value yields a single Y value.

Lemma 6.1 *$X \rightarrow Y$ holds iff $\mathcal{H}_{X \rightarrow Y} = 0$.*

PROOF $\Rightarrow$: Recasting this in terms of probabilities, given any $x_i \in X$, there is a single $y_j \in Y$ such that $p_{i,j} > 0$, and consequently $p_{i,j} = p_i$, and $\mathcal{H}(Y|X = x_i) = 0$ for any x_i.

$\Leftarrow$: Since $\mathcal{H}_{XY} = \mathcal{H}_X$, $\mathcal{H}(Y|x_i) = 0$, for any x_i. Further, $p_i > 0$ for all i. By **Remark** 1.1, $Y|x_i$ is a singleton; hence, $X \rightarrow Y$.

6.2 Armstrong's axioms

Armstrong's axioms [8] are important for functional dependency theory because they provide the basis for a dependency inferencing system. There are commonly three rules given as the Armstrong Axioms, which are merely specializations of the above inequalites.

1. **Reflexivity** If $Y \subseteq X$ then $X \rightarrow Y$
2. **Augmentation** $X \rightarrow Y \Rightarrow XZ \rightarrow YZ$
3. **Transitivity** $X \rightarrow Y \ \& \ Y \rightarrow Z \Rightarrow X \rightarrow Z$

Theorem 6.1 The Armstrong axioms can be derived directly from InD inequalities.

PROOF Reflexivity follows directly from **Lm** 5.1, augmentation from **Lm** 5.7, and transitivity from **Lm** 5.8.

An additional three rules derived from the axioms are often cited as fundamental: union, psuedotransitivity, decomposition. These also follow from **Lm** 5.3, **Lm** 5.11, and **Lm** 5.10 respectively. Interestingly, a critical distinction between Armstrong's axioms and InD inequalities is that in the former, union can be derived from the original three axioms, whereas the latter union must be derived from first principles.

6.2.1 Fixed arity dependencies

Lemma 6.1 for FDs is alternatively a statement about the number of distinct values any $x_i \in X$ determines (we work through an example to motivate this). In the case of FDs and $X \rightarrow Y$, **count-distinct**$(\pi_Y(\sigma_{X=x_i}(\mathbf{r}))) = 1$, for any $x_i \in X$ in a non-empty $\mathbf{r}$. In practice, however, the size is often not unity and FDs are ill-suited for this *e.g.*, consider a $\{\texttt{Parent}, \texttt{Child}\}$ relation $\mathbf{r}$. Biologically **count-distinct**$(\pi_{\texttt{Parent}}(\sigma_{\texttt{Child}=c}(\mathbf{r}))) = 2$, for any child $c \in \texttt{Child}$. InDs measures can be used to model this dependency easily; $\mathcal{H}_{\texttt{Child} \rightarrow \texttt{Parent}} = \log 2 = 1$.

6.3 Multivalued dependencies

In the following, X, Y, Z partition R. Multivalued dependencies (MVDs) arise naturally in database design and are intimately related to the (natural) join operator $\bowtie$. A multivalued dependency, written $X \twoheadrightarrow Y$, holds if $\mathbf{r} = \pi_{XY}(\mathbf{r}) \bowtie \pi_{XZ}(\mathbf{r})$. Intuitively, we see that the values of Y and Z are not related to each other *wrt* an particular value of X.

Lemma 6.2 *MVD count.* Assume $X \twoheadrightarrow Y$ in $\mathbf{r}$. Then for all x_i, y_j, z_k

$$\begin{aligned}&\mathbf{count}(\sigma_{X=x_i,Y=y_j,Z=z_k}(\mathbf{r}))\\ &= \mathbf{count}(\sigma_{X=x_i,Y=y_j}(\mathbf{r}))\mathbf{count}(\sigma_{X=x_i,Y=y_j}(\mathbf{r}))\end{aligned}$$

PROOF By definition of MVDs.

Lemma 6.3 *$X \twoheadrightarrow Y|Z$ holds iff $\mathcal{H}_{X\to Y} + \mathcal{H}_{X\to Z} = \mathcal{H}_{X\to YZ}$.*

PROOF $\Rightarrow$: By **Lm** 6.2, the conditional probabilities of Y, Z *wrt* X must be independent, which is the condition required in **Lemma** 5.3 for equality to hold.

$\Leftarrow$:By **Lemma** 5.3 for equality, the conditional probabilities of Y, Z *wrt* X are independent; hence, by **Lemma** 6.2, $X \twoheadrightarrow Y$.

Since acyclic join dependencies can be characterized by a set of MVDs, it is clear that InD inequalities can characterize them as well, though the "work" is really done by the characterization of the set of MVDs.

6.4 Additional InD inference rules

There are three standard rules of MVD inference:

1. **Complementation** If $X \twoheadrightarrow Y$, then $X \twoheadrightarrow (R - XY)$
2. **Augmentation** For $V \subseteq W$, if $X \twoheadrightarrow Y$ then $XW \twoheadrightarrow YV$
3. **Transitivity** If $X \twoheadrightarrow Y$ and $Y \twoheadrightarrow Z$, then $X \twoheadrightarrow (Z - Y)$

Both complementation and augmentation trivially true under InD inequalies. The last rule, transitivity, is rather interesting. For its proof, we find an alternative characterization of MVDs. Intuitively, the proof establishes that ...

Lemma 6.4 *$X \twoheadrightarrow Y$ iff $\mathcal{H}_{X\to Z} = \mathcal{H}_{XY\to Z}$*

PROOF

$$\begin{aligned}\mathcal{H}_{X\to Y} + \mathcal{H}_{X\to Z} &= \mathcal{H}_{X\to YZ} && \mathbf{Lm}\ 6.3\\ \mathcal{H}_{X\to Y} + \mathcal{H}_{X\to Z} &= \mathcal{H}_{X\to Y} + \mathcal{H}_{XY\to Z} && \mathbf{Lm}\ 5.4\\ \mathcal{H}_{X\to Z} &= \mathcal{H}_{XY\to Z}\end{aligned}$$

Interestingly, this is an alternative characterization of MVDs. In this case, Y does not contribute any information about Z.

Lemma 6.5 $\mathcal{H}_{X\to VW} - \mathcal{H}_{XY\to WV} \geq \mathcal{H}_{X\to V} - \mathcal{H}_{XY\to V}$.

PROOF

$\mathcal{H}_{XV\to W} \geq \mathcal{H}_{XYV\to W}$ **Lm** 5.5
$\mathcal{H}_{XV\to W} - \mathcal{H}_{XYV\to W} +$
$\quad \mathcal{H}_{X\to X} + \mathcal{H}_{XY\to XY} \geq 0$ **Lm** 5.1
$\mathcal{H}_{VWX} - \mathcal{H}_X - \mathcal{H}_{XYWV} + \mathcal{H}_{XY} -$
$\quad \mathcal{H}_{XV} + \mathcal{H}_X + \mathcal{H}_{XYV} - \mathcal{H}_{XY} \geq 0$ **Lm** 4.1
$\mathcal{H}_{X\to VW} - \mathcal{H}_{XY\to VW} \geq$
$\quad \mathcal{H}_{X\to V} - \mathcal{H}_{XY\to V}$ **Lm** 4.1

Lemma 6.6 As a consequence of **Lm** 6.5, $\mathcal{H}_{X\to VW} = \mathcal{H}_{XY\to WV}$, then $\mathcal{H}_{X\to V} = \mathcal{H}_{XY\to V}$.

Lemma 6.7 If $Y \twoheadrightarrow W|VX$, then $XY \twoheadrightarrow W|V$ by **Lm** 5.12.

Lemma 6.8 Let $XYWV = R$. If $X \twoheadrightarrow Y|WV$ and $Y \twoheadrightarrow W|XV$, then $\mathcal{H}_{X\to Y} + \mathcal{H}_{X\to V} + \mathcal{H}_{X\to W} = \mathcal{H}_{X\to R}$

PROOF

$\mathcal{H}_{X\to R}$
$= \mathcal{H}_{X\to Y} + \mathcal{H}_{X\to WV}$
$= \mathcal{H}_{X\to Y} + \mathcal{H}_{XY\to WV}$ **Lm** 6.4
$= \mathcal{H}_{X\to Y} + \mathcal{H}_{XY\to W} + \mathcal{H}_{XY\to V}$ **Lm** 6.7
$= \mathcal{H}_{X\to Y} + \mathcal{H}_{X\to W} + \mathcal{H}_{X\to V}$ **Lms** 6.4, 6.5

Lemma 6.9 *Transitivity for MVDs.*

PROOF

$\mathcal{H}_{X\to R}$
$= \mathcal{H}_{X\to Y} + \mathcal{H}_{X\to W} + \mathcal{H}_{X\to V}$**Lm** 6.8
$\geq \mathcal{H}_{X\to W} + \mathcal{H}_{X\to YV}$**Lm** 5.3
$\geq \mathcal{H}_{X\to WYV} = \mathcal{H}_{X\to R}$**Lm** 5.3

6.5 Rules involving both FDs and MVDs

There are a pair of rules that allow mixing of FDs and MVDs:

1. **Conversion** $X \to Y \Rightarrow X \twoheadrightarrow Y$
2. **Interaction** $X \twoheadrightarrow Y$ & $XY \to Z \Rightarrow X \to Z$

The rule for conversion is trivial. Interaction follows from **Lm** 6.4.

In Section 6.2, we stated a critical difference between Armstrong axioms and InD inequalities was the distinction between what were axioms and derivable rules. Additionally, there appear to be other fundamental differences between FDs and MVDs, and InD inequalities. For example, consider the following problem. Let R be a schema and $F = \{X \to Y | X, Y \subseteq R\}$ a set of FDs over R. Let $\mathcal{I}(R, F)$ be the set of all relation instances over R that satisfy F. For $X \subseteq R$, let $\Pi_X(\mathcal{I}(R, F)) = \{\pi_X(\mathbf{r}) | \mathbf{r} \in \mathcal{I}(R, F)\}$. The question is whether there exists a set G of FDs over X such that $\Pi_X(\mathcal{I}(R, F)) = \mathcal{I}(X, G)$. It is known that in general such a G does not exist. Further, a similar negative result holds for MVDs. InD measures are a broader class than FDs and MVDs, and the expectation is that a theorem holds: it does, trivially since all relation instances satisfy any set of InD inequalities.

7 InD measure constraints

To summarize the previous sections, we have defined InD measures on an instance, values that reflect how much information is additionally required about a second set of attributes given a first set. We have proved a number of arithmetic equalities and inequalities between various InD measures for a given schema; these (in)equalities must hold for any instance of that schema. And we have shown that constraining certain InD measures, or simple expressions involving InD measures, to 0 imposes functional or multivalued dependences on the instances. We now generalize this last step by considering arbitrary numeric constraints upon InD measures, *e.g.*, $\mathcal{H}_{X \to Y} \geq 4/9$. A relation instance $\mathbf{r}$ over $R \supseteq \{X, Y\}$ is a solution to this constraint if $\mathcal{H}^{\mathbf{r}}_{X \to Y} \geq 4/9$ by standard arithmetic. Formally,

Definition 7.1 An *InD constraint system* over schema R is an $m \times n$ linear system

$$\begin{array}{rcl} a_{11}\mathcal{H}_{X_1} + a_{12}\mathcal{H}_{X_2} + \ldots + a_{1n}\mathcal{H}_{X_n} & \geq & b_1 \\ a_{21}\mathcal{H}_{X_1} + a_{22}\mathcal{H}_{X_2} + \ldots + a_{2n}\mathcal{H}_{X_n} & \geq & b_2 \\ & \vdots & \\ a_{m1}\mathcal{H}_{X_1} + a_{12}\mathcal{H}_{X_2} + \ldots + a_{mn}\mathcal{H}_{X_n} & \geq & b_m \end{array}$$

where $X_i \in 2^R$, $a_{ij}, b_i \in \mathbb{Q}$.

The constraint system is characterized by $\mathbf{A} = [a_{ij}]$, $\mathbf{b} = (b_1, \ldots, b_m)$, and $\mathbf{X} = (X_1, \ldots, X_n)$ and will be written as $\mathbf{A}\mathbf{H}_{\mathbf{X}} \geq \mathbf{b}$, where $\mathbf{H}_{\mathbf{X}} = (\mathcal{H}_{X_1}, \ldots, \mathcal{H}_{X_n})^{Transpose}$. Observe that Definition 7.1 is sufficient to describe any InD measure or inequality. InD constraint systems can be as simple as requiring a single FD or as extensive as specifying the entropies of all subsets of R. However, not every $\mathbf{A}$, $\mathbf{b}$, and $\mathbf{X}$ make sense as applied to a relation instance. Either the $\mathbf{A}$ and $\mathbf{b}$ may admit no solutions (*e.g.* $\mathcal{H}_X - \mathcal{H}_Y > 5, \mathcal{H}_Y - \mathcal{H}_X > 7$) or the solutions may violated the InD measure constraints for $\mathbf{X}$ (*e.g.* $\mathcal{H}_{X \to Y} = 3, \mathcal{H}_{Y \to Z} = 1, \mathcal{H}_{X \to Z} = 5$ violates **Lm** 5.8).

Definition 7.2 An InD constraint system $\mathbf{A}$, $\mathbf{b}$, $\mathbf{X}$ is *feasible* provided that the linear system $\mathbf{A}$, $\mathbf{b}$ plus all InD measure constraints inferable from $\mathbf{X}$ is solvable.

Observe that a solution to this extended system involves finding values for each $\mathcal{H}_{X_1}, \ldots, \mathcal{H}_{X_n}$.

7.1 Instances for feasible constraint systems

The question naturally arises whether an instance always exists for a feasible constraint system. The affirmative answer to this question, whose proof is sketched below, provides InD measures with an analog to completeness.

Before venturing into the proof of the theorem itself, we prove a simple result merely for the sake of providing intuition for what comes after. There are two things to be observed while reading the following proof: first, the duality between instance counts and approximate probabilities, and, second, the way interpolation occurs.

Lemma 7.1 Given a rational $c \geq 0$, there exists a relation instance $\mathbf{r}$ over a single attribute A such that $|\mathcal{H}^{\mathbf{r}}_A - c| < \epsilon$ for any $0 < \epsilon$.

PROOF Let $k = \lfloor 2^c \rfloor$ and $f(x) = k\mathcal{H}(1/(k+x)) + \mathcal{H}(x/(k+x))$ for $0 \leq x \leq 1$. Then $f(0) = \log k \leq c \leq f(1) = \log(k+1)$. By the intermediate value theorem, since f is a continuous function on the interval $[0,1]$, and c is a value between $f(0)$ and $f(1)$, then there exists some $a \in [0,1]$ such that $f(a) = c$. Then $p_i = 1/(k+a)$, for $1 \leq i \leq k$ and $p_{k+1} = a/(k+a)$ is the probability distribution. From this distribution we can approximate $\mathbf{r}$ by constructing an instance $\hat{\mathbf{r}}$ over $\{A\}$ with $\langle 1 \rangle, \ldots, \langle k+1 \rangle$ distinct values that is sufficiently large such that if $\mathbf{count}(\sigma_{A=i}(\hat{\mathbf{r}})) = \lfloor \mathbf{count}(\hat{\mathbf{r}}) \cdot p_i \rfloor$, then $|\mathbf{count}(\sigma_{A=i}(\hat{\mathbf{r}}))/\mathbf{count}(\hat{\mathbf{r}}) - p_i| < \epsilon$.

While this proof is non–constructive, we can find a suitable x by, for example, binary search.

Theorem 7.1 *Instance existence.* For any feasible constraint system $\mathbf{A}$, $\mathbf{b}$, and $\mathbf{X}$, and any $\epsilon > 0$, there is a relation instance $\mathbf{r}$ that satisfies $\mathbf{A}$, $\mathbf{b}$, and $\mathbf{X}$ within ϵ.

1. Using the observation from Definition 7.2, solve $\mathbf{A}$, $\mathbf{b}$, and $\mathbf{X}$ for fixed values for $\mathcal{H}_{A_1}, \ldots$.
2. Pick $m > 1/\epsilon$
3. Give every attribute a value with large probability, namely $1 - (1/2m)^k$, where k is the number of attributes. Note that these highly probable attributes contribute a negligible amount to any entropy since their probabilities are so close to 1.
4. The remaining probabilities for each attribute A_i will be divided among b_i equal size buckets. Thus, $\mathcal{H}_{A_i} = (1/m^k)(\log(1/(m^k \times b_i)) \simeq 1/m^k \times \log b_i$. Find b_i such that $\alpha_i - 1/m^k \log(b_i) < \epsilon$

   Remark 2 *Wlog*, the A_i are ordered in decreasing entropy. Hence $b_i \geq b_{i+1}$.

   We will add attributes in order $A_1, A_2, \ldots,$
5. At stage $i+1$, construction has included $A_1, \ldots, A_i$, and we are adding A_{i+1}; that is, we already have $p_{j_1,\ldots,j_i}$ and want to construct $p_{j_1,\ldots,j_{i+1}}$. We also have a single distribution q corresponding to A_{i+1}. We actually construct two distributions $p\ell$ and pu, for "p lower" and "p upper".
   (a) The upper case is simple: A_{i+1} is independent from $A_1, \ldots, A_i$: $pu_{j_1,\ldots,j_i,j_{i+1}} = p_{j_1,\ldots,j_i} \times q_{j_{i+1}}$
   (b) The lower case is found by allocating the q_j among the various p's. Because $b_i \geq b_{i+1}$, there are more than enough i buckets to go around. With some small error, each non-zero p will correspond to a unique $q \neq 0$.
   Error $\mathcal{H}^{p\ell}_{A_i A_{i+1}} - \mathcal{H}^{p\ell}_{A_i} < \epsilon^k$ and by induction $\mathcal{H}^{p\ell}_{A_n A_{i+1}} - \mathcal{H}^{p\ell}_{A_n} < \epsilon^{k-m}$, for $1 \leq m \leq i$.
   Interpolate between p_ℓ and p_u to match other entropies

This is conceptually similar to **Lm** 7.1, but relies upon the unusual structure of pu caused by the almost–unity cases of p and q and another iteration.

8 Applications and extensions

We have presented a formal foundation incorporating information theory in relational databases. There are many interesting and valuable applications and extensions of this work that we are already pursuing.

8.1 Datamining

Datamining [3], the search for interesting patterns in large databases, motivated our initial work, our interest in establishing what it means to be "interesting." A primary objective here is to certainly find all the InD measures $\mathcal{H}_{X\to Y} \leq \delta$ given an instance $\mathbf{r}$ over R. The search in $\mathbf{r}$ takes place upon the lattice of $\langle 2^R, \subseteq\rangle$, where $\mathcal{H}_{X\to Y} \leq \delta$ is checked for every $X \subsetneq Y$. The InD inequalities facilitate this search.

Kivinen *et. al.* [4], considers finding approximate FDs. The central notion is that of *violating pair*; for an instance $\mathbf{r}$ over R and $X, Y \subseteq R$, a pair of tuples $s, t \in \mathbf{r}$ *violates* $X \to Y$ if $s.X = t.X \Rightarrow s.Y \neq t.Y$. They define three normalized measures g_1, g_2, g_3 are based upon the number of violating pairs, the number of violating tuples, and the number of violating tuples removed to achieve a dependency, respectively. The authors state that problematically the measures give very different values for some particular relations, and therefore, choosing which measure is the best–if any are–is difficult. We feel that the InD measure can shed some light upon the metrics. The connection between these measures and InD measures is illustrated with three instances $\mathbf{r} = \{\langle \mathtt{a}, 1\rangle, \langle \mathtt{a}, 2\rangle, \langle \mathtt{b}, 1\rangle, \langle \mathtt{c}, 1\rangle, \langle \mathtt{c}, 2\rangle\}$, $\mathbf{s} = \mathbf{r} - \{\langle \mathtt{c}, 2\rangle\} \cup \{\langle \mathtt{a}, 3\rangle\}$ and $\mathbf{t} = \mathbf{s} \cup \{\langle \mathtt{a}, 4\rangle, \langle \mathtt{a}, 5\rangle, \langle \mathtt{a}, 6\rangle, \langle \mathtt{d}, 1\rangle, \langle \mathtt{d}, 1\rangle\}$

	$\mathcal{H}_X$	$\mathcal{H}_{X\to Y}$	g_1	g_2	g_3
$\mathbf{r}$	1.52	.80	.16	.8	.4
$\mathbf{s}$	1.37	.95	.36	.8	.4
$\mathbf{t}$	1.77	1.55	.36	.8	.4

This example shows that $\mathcal{H}_{X\to Y}$ can sometimes make finer distinctions than g_is. On the applications side, Kivinen *et. al* have done substantial work related to approximate FDs as in [4]. The paper is important not only for the notion of approximate dependency, but also a brief discussion about how the errors can be cast into Armstrong Axiom-like inequalities.

8.2 Other Metrics

Rather than considering what information X lacks about Y, we may look at the information X contains about Y, that is $\hat{\mathcal{I}}_{X\to Y} = \mathcal{H}_Y - \mathcal{H}_{X\to Y}$ and its normalized form $\mathcal{I}_{X\to Y} = \hat{\mathcal{I}}/\mathcal{H}_Y$. Some interesting results about $\mathcal{I}$ and $\hat{\mathcal{I}}$ are $\mathbf{max}(\mathcal{I}_{X\to Y}, \mathcal{I}_{X\to Z}) \geq \mathcal{I}_{X\to YZ} \geq \mathbf{min}(\mathcal{I}_{X\to Y}, \mathcal{I}_{X\to Z})$; $0 \leq \mathcal{I}_{X\to Y} \leq 1$; $\hat{\mathcal{I}}_{X\to Y} = \hat{\mathcal{I}}_{Y\to X}$. While $\mathcal{I}$ makes the specification of FDs more natural ($X \to Y$ iff $\mathcal{I}_{X\to Y} = 1$), it cannot be used to characterize MVDs. Another interesting measure that uses additional notions from information theory is *rate* of the language $s = \mathcal{H}^{\mathbf{r}}_X/\mathbf{count}(\mathbf{r})$ which is the average number of bits required for each tuple projected on X. The absolute rate is $s_{ab} = \log(\mathbf{count}(\mathbf{r}))$. The difference $s_{ab} - s$ indicates the redundancy. As X approaches R, the average tuple entropy increases, reducing redundancy. This is pertinant especially to the following section.

8.3 Connections to relational algebra

Examining how InDs behave with relational operators. For example,

Lemma 8.1 Let $R = \{X, Y, Z\}$ and $\mathbf{r}$ be an instance of R. if $\mathbf{r}' = \pi_{XY}(\mathbf{r}) \bowtie \pi_{XZ}(\mathbf{r})$, then $\mathcal{H}_{YZ}{}^{\mathbf{r}} = \mathcal{H}_Y{}^{\mathbf{r}'} + \mathcal{H}_Z{}^{\mathbf{r}'}$.

For instance, when employing a lossless decomposition, how will both the InD measures and rates (from above) change to indicate the decomposition was indeed lossless.

9 Related work

There is a dearth of literature in this area, marrying information theory to information systems. The closest work seems to be Piatesky-Shapiro in [2] who proposes a generalization of functional dependencies, called *probabilitistic dependency* (*pdep*). The author begins with the $pdep1(X) = \sum_i^\ell p_i^2$ (using our notation). To relate two sets of attributes X, Y, $pdep(X,Y) = \sum_i^\ell p_i \sum_j^m p_{ij}^2$. Observe that *pdep* approaches 1 as X comes closer to functionally determining Y. Since *pdep* is itself inadequate, the author normalizes it using proportion in variation, resulting in the known statistical measure $\tau(X,Y) = (pdep(X,Y) - pdep1(Y))/(1 - pdep1(Y))$. If $\tau(X,Y) > \tau(Y,X)$, then $X \rightarrow Y$ is a better FD than $Y \rightarrow X$ (and vice versa). The author describes the expectation of both *pdep* effeciently sample for these values.

In the area of artificial intelligence, an algorithm developed to create decision trees, a means of classification, by Quinlan, notably ID3 [5] and C4.5 [6] uses entropy to dictate how the building should proceed. In this case of supervised learning, an attribute A is selected as the target, and the remaining attributes $R - \{A\}$ the classifier. The algorithm works by progressively selecting attributes from the intial set $R - \{A\}$, measuring be classified properly.

10 Acknowledgements

The authors would like to thank Dennis Groth, Dirk Van Gucht, Chris Giannella, Richard Martin, and C.M. Rood for their helpful suggestions.

References

[1] Bartle, R. G. *The Elements of Real Analysis Second Edition.* John Wiley & Sons, Inc., New York, New York, 1976.

[2] G. Piatetsky-Shapiro. Probabilistic data dependencies. In *Machine Discovery Workshop (Aberdeen, Scotland)* (1992).

[3] G. Piatetsky-Shapiro, U. Fayyad, and P. Smith, Eds. *From data mining to knowledge discovery: An overview.* AAAI/MIT Press, 1996.

[4] Jyrki Kivinen, and Heikki Mannila. Approximate inference of functional dependencies from relations. *Theoretical Computer Science 149* (1995), 129–149.

[5] Quinlan, J. R. Induction of decision trees. *Machine Learning 1*, 1 (1986), 81–106.

[6] Quinlan, J. R. *C4.5: Programs for Machine Learning.* Morgan Kaufman, San Mateo, CA, 1993.

[7] Roman, S. *Coding and Information Theory.* Springer-Verlag, New York, New York, 1992.

[8] Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of Databases.* Addison-Wesley Publishing Company, New York, New York, 1995.

[9] Thomas Cover, and Joy Thomas. *Elements of Information Theory.* John Wiley & Sons, Inc., New York, New York, 1991.

[10] Ullman, J. D. *Principles of Database and Knowledge-Base Systems Vol. 1.* Computer Science Press, Rockville, Maryland, 1988.