LIBERTAS MATHEMATICA , VOL. 4 , 1984

OUTER MEASURES ON FINITE SETS AND INTEGRITY CONSTRAINTS IN RELATIONAL DATABASES

Dan A. Simovici and Colin Godfrey

1. Introduction . This note establishes a link between some aspects of measure theory and two basic forms of integrity constraints in relational databases: functional and multivalued dependencies .

We use hereby the standard terminology of measure theory and of relational databases . The reader is referred to [2] and [1], respectively , for the basic facts and notations of these disciplines .

Let $R = A_1 \ldots A_n$ be a relation scheme on the attributes $A_1 , \ldots , A_n$. For a relation r(R) on the scheme R , we shall denote by $proj_K(r)$ the projection of r onto the set K of attributes (with duplicate tuples removed !) . Here , $K \subseteq R$.

If $X \subseteq R$ and x is an X-value , the selection of the relation r giving all tuples having their X-component equal to x will be denoted by $sel_{X=x}(r)$.

Consider the mapping $\mu_r : PR \to R_+$ given by $\mu_2(K) = \log_2 |proj_K(r)|$. It is clear that $\mu_r(K)$ measures the time complexity of searching the projection of r onto K , for a tuple with a prescribed K-component .

Proposition 1 . For every relation r(R) , μ_r is an outer measure on R .

Proof . Let $\{ Y_i \mid i \in I \} \subseteq PR$ be a collection of subsets of R , and consider the mapping

$$\phi : proj_Y(r) \to \prod_{i \in I} proj_{Y_i}(r) ,$$

where $Y = \{Y_i \mid i \in I\}$, given by $\phi(t) = (\ldots, t(Y_i), \ldots)$. Here t(Z) is the projection of the tuple t on the set of attributes $Z \subseteq R$. It is easy to verify that this mapping is injective , hence

$$|proj_T(r)| \leq \prod_i |proj_{Y_i}(r)| .$$

By applying the logarithms we obtain immediately the subadditivity of μ_r.

We leave to the reader the direct verification of other properties of outer measures .

2. Outer measures and multi-valued dependencies .

The state of μ_r in the X-value x (where $X \subseteq R$) is the mapping $\mu_{x,r}$: $PR \to R$ defined by $\mu_{x,r}(K) = \mu_{sel_{X=x}(r)}(K)$ for all $K \in PR$.

Since $sel_{X=x}(r)$ is again an R-relation , it is clear that every state of an outer measure generated by a relation is an outer measure with the same character.

The importance of these outer measures is highlighted by

Proposition 2 . The relation r(R) , on the relational scheme R , satisfies the multivalued dependency $X \twoheadrightarrow Y$ if Y is $\mu_{x,r}$ - measurable for every X-value x .

Proof . Suppose that r satisfies the multivalued dependency $X \twoheadrightarrow Y$, that is r admits a lossless decomposition into $proj_{XY}(r)$ and $proj_{X\bar{Y}}(r)$.

We shall remark that the injection

$$\phi : proj_K(sel_{X=x}) \to proj_{K\cap Y}(sel_{X=x}(r)) \times proj_{K\cap \bar{Y}}(sel_{X=x}(r))$$

is also surjective for every $K \in PR$.

Indeed , let $t_1 \in proj_{K\cap Y}(sel_{X=x}(r))$ and $t_2 \in proj_{K\cap \bar{Y}}(sel_{X=x}(r))$. There are t_1' , $t_2' \in sel_{X=x}(r)$, such that $t_1 = t_1'(K\cap Y)$ and $t_2 = t_2'(K\cap \bar{Y})$.

Due to the multivalued dependency $X \twoheadrightarrow Y$, there are u , $v \in sel_{X=x}(r)$, such that $u(Y) = t_1'(Y)$, $u(\bar{X}\cap\bar{Y}) = t_2'(\bar{X}\cap\bar{Y})$, and $v(Y) = t_2'(Y)$, $v(X\cap Y) = t_1'(\bar{X}\cap\bar{Y})$. These relations show that $t_1 = proj_{K\cap Y}(u)$, and $t_2 = proj_{K\cap\bar{Y}}(u)$, that is $\phi(u) = (t_1 , t_2)$. Thus , ϕ is surjective , hence $\mu_{x,r}(K) = \mu_{x,r}(K\cap\bar{Y})$, which proves that Y is $\mu_{x,r}$- measurable for all X-values x .

Conversely , suppose that Y is $\mu_{x,r}$- measurable . In this case

$$\mu_{x,r}(R) = \mu_{x,r}(R\cap Y) + \mu(R\cap\bar{Y}) ,$$

that is

$$|sel_{X=x}(r)| = |proj_Y(sel_{X=x}(r))| \cdot |proj_{\bar{Y}}(sel_{X=x}(r))|$$

for all X-values x , which is a well known criterion for the existence of the multivalued dependency $X \twoheadrightarrow Y$ (see [1]) .

Using the states of outer measures , it is possible to express the exis - tence of functional dependencies satisfied by relations .

Proposition 3 . The relation r(R) satisfies the functional dependency $X \to Y$ for X , $Y \in PR$ if and only if $\mu_{x,r} = 0$ for any X-value x .

Proof . If r satisfies $X \to Y$, then any X-value x uniquely determines the Y-value of a tuple t of r . Thus , $proj_Y(sel_{X=x}(r))$ consists of a single element and this implies $\mu_{x,r}(Y) = 0$. The converse implication is immediate .

In view of the fact that every set S for which $\mu(S) = 0$ is μ - measurable (where μ is an outer measure) , it is possible to get a better image of the role played by functional dependencies with respect to multivalued dependencies , by using the tools offered by measure theory .

3. Partitions of finite sets and outer measures .

Let M be a finite set and assume that $\pi = \{ B_i \mid i \in I\}$ is a partition of M having the blocks B_i .

A set $P \subseteq M$ is π - saturated if there is $J \subseteq I$, such that $P = \cup \{B_i \mid i \in J \}$. It is easy to see that the family S_π of π - saturated sets is a subalgebra of the Boolean algebra (PM , M , $\emptyset$, U , ,) of all subsets . If $m: S_\pi \to R_+$ is a measure , then m will generate an outer measure $\mu : PM \to R_+$, via the standard construction

$$\mu(P) = \Sigma \{ m(B_i) \mid B_i \cap P \neq \emptyset\}$$

The class of μ - measurable sets equals S_π . Moreover , we notice that m is fully defined by its restriction m_o to the subalgebra S_π (that is , to the blocks of the partition π) .

Proposition 4 . For every regular outer measure $\mu : PR \to N$, there is a relation r(R) on the relational scheme R , such that $\mu_r = \mu$.

Proof . Let MEAS(μ) be the subalgebra of all μ - measurable sets . The minimal set of this collection form a partition π_μ of the set R . Let us assume that $\pi_\mu = \{ B_1,\ldots,B_r \}$ and let m_o be the restriction of μ to these atoms .

Clearly , m_o generates a measure on S_{π_μ} , and the outer measure generated by m via the standard mechanism aforementioned coincides with μ . We shall consider a table T_o having r columns , one for each block of $\pi_\mu = \{B_1,\ldots,B_r\}$. T_o will contain 2^σ rows , where $\sigma = \Sigma \{\mu(B_i) \mid 1 \leq i \leq r\}$. The j-th row will contain a sequence of σ 0's and 1's , representing the binary equivalent of a number j , where $0 \leq j \leq 2^\sigma - 1$; $\mu(B_i)$ consecutive bits will be placed in the column corresponding to B_i . If $\mu(B_i) = 0$, the whole B_i column will contain only 0's .

Example . Suppose that $R = A_1A_2A_3A_4$, $\pi_\mu = \{A_1A_2 , A_3A_4\}$, $\mu(A_1A_2) = 2$ and $\mu(A_3A_4) = 1$; let $B_1 = A_1A_2$ and $B_2 = A_3A_4$. We shall have

Table T_o

B_1		B_2
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

Table T_o'

B_1	B_2
0	0
0	1
1	0
1	1
2	0
2	1
3	0
3	1

In T_o we shall replace the binary sequences by their corresponding numerical equivalents ; the table thus obtained is denoted T_o' .

If $A \in B_i$, the column corresponding to A in the final table T (containing the desired relation r) will be a duplicate of the column B_i .

If $K \subseteq R$, $K = A_{i_1}A_{i_2}\ldots A_{i_n}$, $\text{proj}_K(r)$ will contain $2^{\mu(A_{i_1})}2^{\mu(A_{i_2})}\ldots 2^{\mu(A_{i_n})}$ distinct values . Thus ,

$$\mu_r(K) = \log_2(2^{\mu(B_{i_1})}2^{\mu(B_{i_2})}\ldots 2^{\mu(B_{i_n})})$$

where $A_{i_p} \in B_{i_p}$ for $1 \leq p \leq n$. Therefore ,

$$\mu_r(K) = \Sigma \{\mu(B_{i_p}) \mid 1 \leq p \leq n\} = \mu(K),$$

which concludes the proof , since $\{B_{i_1}, \ldots, B_{i_n}\}$ is exactly the set of blocks intersecting K .

Example . The final table corresponds to the measure μ considered in the previous example is given below .

A_1	A_2	A_3	A_4
0	0	0	0
0	0	1	1
1	1	0	0
1	1	1	1
2	2	0	0
2	2	1	1
3	3	0	0
3	3	1	1

For a relation r(R) on a scheme R , the set $X \subseteq R$ is $\mu_{x,r}$ - measurable for any X-value x , since $\mu_{x,r}(X) = 0$. Thus , the atoms of a subalgebra MEAS($\mu_{x,r}$) included in X are all of $\mu_{x,r}$-measure 0 .

Let W_x be the largest element of measure 0 of $\mu_{x,r}$, in other words

$$W_x = \cup \{Y \mid Y \subseteq R , \ \mu_{x,r}(Y) = 0\} .$$

The closure of X with respect to the functional dependencies satisfied by the relation r is $X^+ = W_x$ x is an X-value . Therefore , if for any X-value x we have $W_x = X$, the relation r satisfies only trivial functional dependencies .

REFERENCES

[1] Maier , D. , *The Theory of Relational Databases* . Computer Science Press , Rockville , Maryland , 1983 .

[2] Munroe , M. E. , *Introduction to Measure and Integration* . Addison - Wesley Publishing Co. , Reading , Mass. , 1953 .