Lecture 17

Plan

find bounds for estimators

achieve [illegible] guarantees

distinct elements

$P\{|X - \text{true Val}| > \varepsilon\} < \delta$

Sampling

$$X = \alpha \sum_{t \in I} X_t \cdot \delta(t)$$

Sketching

$$X = \sum_{t \in I} \xi_t = \sum_{i \in I} \xi_i f_i$$

$$Z = X^2$$

Size of Join Estimation

R, S two relations

A — common attribute

$$|R \bowtie S| = \sum_i f_i g_i = \langle f, g \rangle$$

f_i — freq of i in R

g_i — — " — S

$$X_R = \sum_{t \in R} \xi_t = \sum_{i \in D} f_i \xi_i$$

$$X_S = \sum_{t \in S} \xi_t = \sum_{i \in D} g_i \xi_i$$

$$X = X_R X_S$$

$$E[X] = \sum_i \sum_j f_i g_j E[\xi_i \xi_j] = \sum_i f_i g_i$$

$$1 \quad \xi \text{ indep of } \xi'$$

$$E[w] = \sum_{i,j} f_i g_j E[\xi_i \xi_j] = 0$$

$$\xi = \xi'$$

$E(X^2) = E[X_n^2]$

$\langle x,x\rangle \langle y,y\rangle \geq \langle x,y\rangle^2$

$\left(\sum_i y_i\right)^2 + \left(\sum_i S_i\right)^2 - 2\sum_i S_i^2$

$E(X^2) - E(X)^2 \leq 2 \ldots (R)\ldots(S)$

Possible generalization

1° Generalizing natural joins

$P(A, B) \quad S(A)$

$$\sum_{t \in P \bowtie S} t_B = \sum_i \hat{f}_i \cdot g_i$$

$$\forall_i \quad \sum_{t \in R} t_B = \sum_{i \in D} i \left(\sum_{\substack{t \in S \\ t_A = i}} t_B \right)$$

$$\forall_i = \sum_{i \in D} g_i f_i$$

$$\sum_{t \in R \bowtie S} t_B = \sum_{i \in I} \cdot \left(\sum_{\substack{t \in R \bowtie S \\ t_A = i}} t_B \right)$$

$$\underbrace{S_i \sum_{\substack{t \in R \\ t_A = i}} t_B}$$

2° Multiple relations

$\underbrace{R, S}_{A}, \underbrace{T}_{B}$