Lecture 8

Sampling

$$X_y = \begin{cases} 1 & \text{if } y \in D' \\ 0 & \text{oth} \end{cases}$$

$$E(X_y) = \frac{|D'|}{|D|}$$

sampling without replacement

$$\delta_{y,z} = \begin{cases} 1 & y = z \\ 0 & \text{oth} \end{cases}$$

Q: what does $y = z$ mean?

1° $y = z$ numerically $y = z$, $y, z \in \mathbb{R}$ modulo noise

2° $y = z$ if they refer to the same piece of D

How to assign identities

- pick any order
- index in order is id

val	1	1	2	3	4	4
id	1	2	3	4	5	6

$E(X_y X_z) =$

$$\sum_{y \in D} \sum_{z \in D} \left(\delta_{y,z} + \quad \right) X_z$$

From now on

D will have relational structure

$D = \{R_1, \dots, R_k\}$

R_i is a relation with its tuples identified (primary key)

$f(D) = f(R_1, \dots, R_k)$

A	B	ID
1	2	1
1	2	2
3	4	3

$$\int_{t,t'} = \begin{cases} 1 & \text{if } t \text{ is the same as } t' \\ 0 & \text{otherwise} \end{cases}$$

Previous strategies

$D = \{R\}$

$R(A)$

$$f(D) = \sum_{t \in R} t.A$$

$$X = f(D) = f(D') = \alpha \sum_{t \in R'} t.A$$

$D' = \{R'\}$ $R'(A)$

R' is a sample of R taken without replacement

$\alpha = \frac{|R|}{|R'|}$

$$f(D) = \sum_{t \in D} t.A$$

$$X = \alpha \cdot \sum_{t \in R} X_t \cdot t.A$$

$$X_t = \begin{cases} 1 & \text{if } t \in R' \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \alpha \sum_{t \in R} E(X_t) \cdot t.A$$

$$E[X^2] = \alpha^2 \sum_{t \in R} \sum_{i \in R} E(X_t X_i) \cdot t.A \cdot i.A$$

Problem 2

$D = \{R(A)\}$

$$f(D) = \sum_{t \in R} g(t.A)$$

where $g(x)$ is some fct

$\forall t \in R, \; X_t = \begin{cases} 1 & \text{if } t \in R' \\ 0 & \text{oth.} \end{cases}$

$$E[X_t] = \frac{|R'|}{|R|} = \frac{1}{\alpha}$$

$t, t' \in R$

$$E[X_t X_{t'}] = P[t \in R' \wedge t' \in R'] = P(t \in R') \cdot P(t' \in R' \mid t \in R') = \frac{1}{\alpha} \begin{cases} \frac{1}{\alpha_1} & \text{if } t' \neq t \\ 1 & \text{oth} \end{cases}$$

$$\alpha_1 = \frac{|R'|-1}{|R|-1}$$

$$E[X_t X_{t'}] = \frac{1}{\alpha}\left(\delta_{t,t'} + (1-\delta_{t,t'})\frac{1}{\alpha_1}\right) = \frac{1}{\alpha}\left(\frac{1}{\alpha_1} + \delta_{t,t'}\left(1 - \frac{1}{\alpha_1}\right)\right)$$

X_t + prop (have nothing to do with $g(\cdot)$)

def $$X = \alpha \sum_{t \in R'} g(t.A) = \alpha \sum_{t \in R} X_t \, g(t.A)$$

$$E[X] = \alpha \sum_{t \in R} E[X_t] \cdot g(t.A) = \sum_{t \in R} g(t.A) = f(D)$$

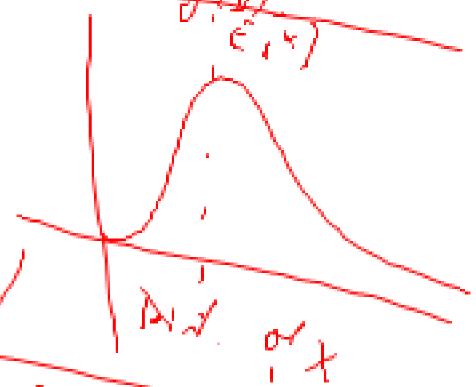
($E[X_t] = \frac{1}{\alpha}$)

Example:

g encodes selectivity predicate:

$g(A) = \begin{cases} A & \text{if } A \geq 10 \\ 0 & \text{oth.} \end{cases}$

SELECT SUM A FROM R WHERE A ≥ 10



$$E[X^2] = \alpha^2 \sum_{t \in R} \sum_{t' \in R} E[X_t X_{t'}] \cdot g(t.A)\, g(t'.A)$$

$$= \alpha^2 \frac{1}{\alpha \alpha_1} \left(\sum_{t \in R} g(t.A)\right)^2 + \alpha^2 \frac{1}{\alpha}\left(1 - \frac{1}{\alpha_1}\right) \sum_{t \in R} g(t.A)^2$$