CSI T5300: Advanced Database Systems

E10: Exercises on Selectivity Estimation

Dr. Kenneth LEUNG

Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Hong Kong SAR, China
Consider relation R(A, B, C). Assume:
- R contains 10000 records.
- A has 50 distinct values in the range 1..50
- B has 100 distinct values in the range 1..100

Estimate the sizes of the following operations assuming uniform distribution and independent conditions:

1. $\sigma_{A=10} \ R$
2. $\sigma_{A=10 \land 20<B} \ R$
3. $\sigma_{C=1} \ R$
4. $\sigma_{C=10 \land A=10} \ R$
5. $\sigma_{C=10 \land A=10 \land 20<B} \ R$
1. $\sigma_{A=10} R$
   - $10000 \times (1 / 50) = 200$

2. $\sigma_{A=10 \land 20<B} R$
   - Condition $A=10$, selectivity = $1 / 50$
   - Condition $20 < B$, selectivity = $80 / 100$
   - Overall selectivity = $(1 / 50) \times (80 / 100) = 8 / 500$
   - Estimated result size = 160

3. $\sigma_{C=1} R$
   - 1 (note that $C$ is the primary key)
4. $\sigma_{C=10 \land A=10} R$
   - Condition $C=10$, selectivity = $1 / 10000$
   - Condition $A = 10$, selectivity = $1 / 50$
   - Overall selectivity = $1 / 10000 \times 1 / 50 = 1 / 500000$
   - Estimated result size = $10000 / 500000 = 0.02$

5. $\sigma_{C=10 \land A=10 \land 20<B} R$
   - Condition $C=10$, selectivity = $1 / 10000$
   - Condition $A = 10$, selectivity = $1 / 50$
   - Condition $20 < B$, selectivity = $80 / 100$
   - Overall selectivity = $1 / 10000 \times 1 / 50 \times 20 / 100 = 8 / 5000000$
   - Estimated result size = $0.016$