

Homework 1
Approximate Query Processing
Spring 2006
Deadline: Feb. 20 2006

Alin Dobra

February 8, 2006

In class we considered only the case in which the sample used for estimation is of fixed, predefined size and the sampling is performed without replacement. In this homework, you have to consider two other types of sampling

1. Sample with replacement (i.e. an item can be selected multiple times)
2. Binomial sampling. In binomial sampling, a coin with probability of success p (p is a given parameter) is tossed independently for each tuple in a relation. If the outcome is success, the tuple is included in the sample (this is sampling without replacement as well).

For each of the two types of sampling provide the analysis both for the problem of computing sum type of aggregates for one relation and the generalization to k relations. What is different from the analysis in class in these circumstances.

You can work in groups but you have to write the answer to the problem individually.