Communication-Efficient Probabilistic Algorithms:
Selection, Sampling, and Checking
Corrections and Remarks

Lorenz Hübschle-Schneider

15th July 2021

In this document, I collect corrections and remarks regarding my dissertation [2].

Corrections

Chapter 3.3.1, Improved Sequential Alias Tables, page 60, line 21. Ares Lagae was kind enough to point out that we are not the first to figure out Michael Vose's hint about linear extra space not being necessary. In fact, Galerne, Lagae, et al. have developed an algorithm with this property [1]. While they do not describe or analyse it in the paper, the supplemental material includes source code of an implementation.

Chapter 3.5, Sampling Without Replacement, page 76, line 23. Downsampling using the exponential clocks technique using the original weights yields incorrect results. We need to use the items' multiplicities in the sample S as the new weights. Remove the last sentence of the paragraph 'Queries' and replace it with the following new paragraph:

We now face the difficulty that while weighted sampling without replacement is defined by the order in which items are sampled, the output-sensitive sampling algorithm for weighted sampling with replacement that we use to find S gives us no such information. However, when we regard S as the sequence of ℓ sampled items, then any permutation thereof is equally likely to be the result of an equivalent sequential sampling with replacement with the same overall result. Thus, the probability of any sampled item e to be the first in a randomly permuted sequence is proportional to its multiplicity in S. Once we remove e, the the same is true for the remaining items, etc. Therefore, we can obtain a subsample with the correct probabilities using weighted sampling without replacement on S with the multiplicities in S as the items' (new) weights. This new problem has expected size $\mathcal{O}(k)$, so we can use the exponential clocks technique of Section 3.1.3 by generating appropriate random variates for each of $|S|$ items and perform a parallel selection for the items that receive rank $\leq k$.

References

[1] Bruno Galerne, Ares Lagae, Sylvain Lefebvre, and George Drettakis. Gabor noise by example. *ACM Transactions on Graphics*, 31(4):73:1–73:9, 2012. `doi:10.1145/2185520.2185569`.

[2] Lorenz Hübschle-Schneider. *Communication-Efficient Probabilistic Algorithms: Selection, Sampling, and Checking*. PhD thesis, Karlsruher Institut für Technologie (KIT), 2020. `doi:10.5445/IR/1000127719`.