A randomized variation on a theorem of Lenstra

David B. Shmoys *

Lenstra (1976) states the following theorem (in a rough translation from the original Dutch):

> "At least one half of the references to the article by the traveling salesman quartet of Little, Murty, Sweeney, and Karel contains at least one error.
>
> J.D.C. LITTLE, K.G. MURTY, D.W. SWEENEY, J. KAREL (1963) An algorithm for the traveling salesman problem. *Operations Res. 11*, 972–989."

Lenstra does not elaborate on the proof technique used to obtain this result, but one might presume, given the state of the art when this work was published, not to mention the title of the thesis, that an enumerative method was employed. Enumerative procedures, especially complete enumeration, are often impractically slow methods, and in this setting there is the added complication of the seemingly exponential increase in the number of papers published.

The purpose of this brief note is to provide a significantly faster procedure to derive such a theorem, based on randomization. Suppose that you would like to prove a Lenstra-type theorem for a new publication X. Rather than checking each published article that contains a citation of X to verify its correctness, suppose that one has the ability to identify one such published article uniformly at random from among the universe U of papers that cite X. Let n denote the number of papers in this universe U. We can use such a "black box oracle" as follows: draw k independent samples by using the oracle, and check each of them to determine if X is cited correctly. Then, depending on the specific values of k and n, one can assert, with a particular success probability that is easy to calculate, a lower bound on the fraction of papers that contains an error. In particular, we have relied on a standard calculation, using the so-called Chernoff-Hoeffding bound in order to obtain the following theorem.

Theorem 1 *With probability at least .95, at least one half of the references to the book by the traveling salesman quartet of Lawler, Lenstra, Rinnooy Kan, and Shmoys contains at least one error.*

E.L. LAWLER, J. LENSTRA, A.H.G. RINNOOY KAN, D.B. SHMOYS (eds.) (1985) *The traveling salesman problem.* Wiley, Chichester.

In the spirit of the work of Lenstra, we will not give the details of the proof of this new theorem. However, Google Scholar provides us with the means for generating the random samples, and we have estimated n to be 2450; furthermore, we could rely on relatively few samples (in our case, 100), since the fraction that contains an error appears to be significantly greater than one half. We have estimated each parameter conservatively, so as to ensure the correctness of the theorem.

* shmoys@cs.cornell.edu. School of Operations Research & Information Engineering and Department of Computer Science, Cornell University, Ithaca, NY 14853. A preliminary version of this appeared in the Proceedings of the 60th Annual Meeting of the FoJKL. Research supported in part by NSF grants CCF-0832782 and CCF-1017688.

Many different errors were observed. Not surprisingly, "travelling" is substituted quite often. Of the 2^4 possible subsets of editors, a significant fraction are listed as the supposed editors in at least one citation. Since the sample contained quite a number of pairs and singletons of authors, it would be interesting to see if the full enumeration would also yield the empty set. The claimed publication date is as early as the previous millennium. Each editor has incorrect initials in some citation. And, with the exception of Lenstra, every editor has his family name misspelled at least once (at least in the random sample examined). One reassuring fact is that in no instance selected was one of any of Lenstra's siblings cited instead.

Many significant open questions remain in this line of research. In particular, will the new traveling salesman quartet fare better, or is their TSP book similarly fated? For this question, there is some preliminary evidence that it will fare similarly, though of the roughly 400 citations found by Google Scholar, many actually cite a preliminary technical report, and hence are not part of the true data set. And more generally, perhaps one can prove that there exist values j, k, and ℓ, such that for each at least j-authored publication on the traveling salesman problem, in which the authors have at least k initials in total, and for which there are at least ℓ distinct citations, such a Lenstra-type theorem holds.

Bibliography

D.L. APPLEGATE, R.E. BIXBY, V. CHVÁTAL, W.J. COOK (2006) *The traveling salesman problem.* Princeton University, Princeton.

E.L. LAWLER, J.K. LENSTRA, A.H.G. RINNOOY KAN, D.B. SHMOYS (eds.) (1985) *The traveling salesman problem.* Wiley, Chichester.

J.K. LENSTRA *Sequencing by enumerative methods.* (1976) Mathematisch Centrum, Amsterdam.