DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

(Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 15, 1992	Final Report 1 Jul 89 - 31 Oct 90

4. TITLE AND SUBTITLE

Probability and Statistics Applied to the Theory of Algorithms

5. FUNDING NUMBERS

DAAL03-89-G-0092

6. AUTHOR(S)

J. Michael Steele

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Princeton University

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ARO 26158.2-MA

11. SUPPLEMENTARY NOTES

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This final report summarizes the contribution of the fourteen articles and two doctoral dissertations that were supported by this grant. The central aim of the work has been to understand the asymptotic behavior of the objective function of problems of classical combinatorial optimization, both in the stochastically modeled cases and in the deterministic worst-case. One of the engaging developments of this investigation has been that there are close parallels in these two problems, despite substantial differences in technique.

In addition to reviewing the main contributions, there is a brief discussion of two articles that responded to targets of opportunity. One of these articles received the Wilcoxon Prize for Best Application Paper in "Technometrics."

14. SUBJECT TERMS

Probability, Algorithms, Traveling Salesman Problem, Minimal Spanning Tree, ACE Algorithm, Minimal Matchings, Worst-Case Analysis

15. NUMBER OF PAGES

9

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Probability and Statistics Applied to the Theory of Algorithms

FINAL REPORT

J. Michael Steele
October 15, 1992

U.S. ARMY RESEARCH OFFICE

Grant Number DAAL03-89-G-0092

Probability and Statistics
Applied to the Theory of Algorithms
J. Michael Steele
Grant Number DAAL03-89-G-0092

I. Nature of the Problems Studied.

The central aim of the problems studied under this grant is to understand when and how probability is useful in the theory of algorithms. Of the thirteen articles cited below which acknowledge support from this grant, the majority are of the nature where one uses probability to study a problem that arises from the theory of combinatorial optimization. The most famous problems of Euclidean combinatorial optimization are perhaps the Euclidean traveling salesman problem, the minimal spanning tree problem, and the minimal matching problem. Probability enters the study of such problems in several ways, but one of the most natural and direct is through the development of stochastic models for the problem inputs. One then uses probability theory to understand as deeply as possible the behavior of the associated objective functions. All but a few of the articles reviewed here take this route; but, ironically, there are exceptions that turn out to have had substantial impact. In particular, it has proved useful to pursue the analogy between worst-case and average-case behaviors. Also, there are two papers that are best viewed as addressing targets of opportunity. Finally, as we note in the last section of the report, work has been done under the subsequent Grant DAAL-03-92-G-0110 that

further contributes to the developments reported on here.

II. Summary of Most Important Results

An Inverse Problem of Computational Statistics

The most acknowledged of the articles that were supported by this grant is the piece with Richard D. De Veaux, *"ACE guided transformation method for the estimation of the coefficient of soil water diffusivity"*, ***Technometrics***, **31**, (1989), 91-98. This article received the Wilcoxon Prize for the outstanding article to appear in ***Technometrics*** in 1989. This article is not in the direct line of most of the work done under this grant. Still, there are natural connections, even though the piece is best understood as a successful response to a target of opportunity.

The theme of the article with De Veaux is a central one of applied statistics -- the choice of a transformation of data. The main twist that we provide is to detail the first application of a systematic *automated* search for such a transformation in the context of an inverse problem, which, in our case, comes from the problem of estimating the coefficients in the basic equation of soil-water diffusion.

Several favorable circumstances came together in the article. First, there is considerable benefit in beginning with serious scientific problem; although, the estimation problem may

sound specialized, there are hundreds of articles on soil-water diffusivity. A second critical contribution to the project was the fortuitous contact at Princeton with soil scientists who were familiar with the relevant scientific literature. Finally, the problem offered a good fit with earlier work since it required computational expertise, comfort with applications of PDE's, and willingness to explore a large number of alternatives before settling on pragmatic choices.

At a minimum, the article with De Veaux showed that the ACE algorithm of Breiman and Stone is a genuinely useful tool of computational statistics. It also made progress toward showing that data analytic thinking which one most often sees in softer social science applications can make a contribution to a hard science topic, like the estimation technology of inverse problems. Finally, the central success of the article is that it provides serious candidate for the method of choice in a widely pursued application area.

Subadditive Euclidean Functionals and Related Work

The majority of the work done under this grant is related in one way or another to the technology of subadditive processes. At the root of this theory is Kingman's subadditive ergodic theorem, and the paper of that title offers one of the shortest proofs of Kingman's famous result. The most telling aspect of my

proof is that it is explicitly algorithmic. Thus, it provides a point of view that is not often seen in the more theoretical parts of probability. For example, Kingman's original proof was in the style of a Hahn-Banach existence argument, even though Kingman pointed out a useful analogy to linear programming in his original article.

The pieces "Efficacy of spacefilling heuristics in Euclidean combinatorial optimization" and "Cost of sequential connection for points in space" engage what one can call the theory of "n points in the unit square." The first of these uses some results from the theory of tube volumes (a basic subject of differential geometry) to study the heuristics that can be based on spacefilling curves. The second paper gives a simple but powerful bound on functionals of the edge lengths of the path obtained by
sequentially inserting points into a tour of points in the square. These results are technical but they provide tools that add to the effectiveness of the theory of subadditive Euclidean functionals, a basic theme of this grant.

The next group of papers that deserve review are those that explore the behavior of the objective function of the classical problems of geometric optimization under the so-called "worst-case". To recall a typical result of this field, we let $T(S)$ denote the length of the shortest tour through the points

$S=\{x_1, x_2, \ldots, x_n\} \subset [0,1]^d$ and set

$$\rho(n) = \max\{T(S) : |S| = n\}.$$

The fact established in ***"Worst Case Growth Rates of Some Classical Problems of Combinatorial Optimization"*** is that $\rho(n)$ is asymptotic to $cn^{(d-1)/d}$ as n goes to infinity. One can see without difficulty that $\rho(n)$ is of order $n^{(d-1)/d}$, but one needs to go rather more deeply into the structure of the TSP in order to get an exact asymptotic relationship.

A similar theme was pursued in the case of minimal matching in ***"Worst Case Matchings in the Unit Cube"***. The form of the main result is quite close to that for the TSP (and related functionals), although the technical demands of the later paper were substantial. The central difficulty originates in the possibility of multiple solutions for the worst-case distributions of points.

Exposition and New Techniques

There are three articles that have substantial expositional content and that also offer some research progress: (1) ***"Seedlings in the Theory of Shortest Paths"***, (2) ***"Probabilistic and Worst Case Analyses of Classical Problems of Combinatorial Optimization in Euclidean Space"***, and (3) ***"Probability and***

Statistics in the Service of Computer Science: Illustrations Using the Assignment Problem". The first of these has a number of partially explored ideas (hence seedlings). The main result of the article is a new martingale proof of a strong tail bound for the TSP. The technique developed there has been used in several subsequent works. The second article of this group is the broadest survey to date on the work in this area, and problems posed there have been engaged by several researchers including M. Talagrand, W.S. Rhee, D. Bertsimas, and K. Alexander. The third article of the group is the most expository, but has been also followed up by several researchers including David Aldous.

Semi-matchings and Connection to Linear Constraints

With "*Euclidean Semi-Matchings of Random Samples*", we return to the central theme of the grant. The main result of that article is a probabilistic limit theorem that recalls the Beardwood-Halton-Hammersley Theorem in the context of matching. The most innovative aspect of the article is that it connects the theory of subadditive Euclidean functionals to the theory of linear programming. This is a powerful connection that has not yet been fully explored.

Two Articles that are a Atypical

Two articles that should be singled out for their atypical nature are *"Certifying smoothness of discrete functions and measuring legitimacy of images," **Journal of Complexity,*** 5, (1989), 261-270, and *"Models for managing secrets," **Management Science,*** 35, (1989), 240-248. The first of these is motivated by one of the simplest questions that can be posed in the important area of automatic target recognition (ATR), and the second explores some very simple probability models that aim to provide insight into the processes by which secrets can be kept, or disclosed.

The initial motivation for *"Models for managing secrets"* came from a line in the novel, ***The Hunt for Red October,*** where a Naval commander said "The likelihood of a secret being blown is proportional to the square of the number of people who are in on it." The aim of the article was to examine the possibility of credible probability models in which the commander's intuition corresponded to an analytical fact. The article received quite positive reviews and was quickly published in ***Management Science***. Invited presentations were given on the article at several universities and the Institute for Defense Analysis.

III. Publications and Technical Reports

"Kingman's subadditive ergodic theorem," *Annales de l'Institute Henri Poincare, 25,* (1989), 93-98.

"Certifying smoothness of discrete functions and measuring legitimacy of images," *Journal of Complexity, 5,* (1989), 261-270.

"Models for managing secrets," *Management Science, 35,* (1989), 240-248.

"ACE guided transformation method for the estimation of the coefficient of soil water diffusivity," (with R. De Veaux), *Technometrics, 31,* (1989), 91-98.

"Efficacy of spacefilling heuristics in Euclidean combinatorial optimization," *Operations Research Letters, 8,* (1989), 237-239.

"Cost of sequential connection for points in space," *Operations Research Letters, 8,* (1989), 137-142.

"Worst-case greedy matchings in the unit cube," with T. L. Snyder, *Networks*, 20, (1990), 779-800.

"Seedlings in the theory of shortest paths," in *Disorder in Physical Systems: A Volume in Honor of J. M. Hammersley (G. Grimmett and D. Welsh, eds.)*, Cambridge University Press, 277-306, London, 1990.

"A randomized data structure for ordered sets," (with J. L. Bentley, F. T. Leighton, M. Lepley, and D. F. Stanat), *Advances in Computing Research*, 5, (1989), 413-428.

"Worst case growth rates of some classical problems of combinatorial optimization," (with T. L. Snyder), *SIAM Journal of Computation*, 18, (1989), 278-287.

"Probabilistic and worst case analyses of classical problems of combinatorial optimization in Euclidean space," *Mathematics of Operations Research*, 15,(1991),11-19.

"Euclidean semi-matchings of random samples," *Mathematical Programming*, 53,192,344-349.

"Probability and statistics in the service of computer science: Illustrations using the assignment problem," Invited Paper for the Proceedings of "FSU Statistics Days", *Communications in Statistics*, (Special Issue: Statistics Days at Florida State University), 19, 1990, 4315-4319.

"LeCam's inequality and Poisson approximation," To appear in *The American Mathematical Monthly*.

IV. Participating Scientific Personnel and Advance Degrees Earned

J. Michael Steele (Principal Investigator)

T. L. Snyder(Associate Professor of Computer Science, Georgetown University). Dr. Snyder received some summer support from this grant as a consultant to the research reported here, and he has contributed as a co-author on several of the referenced articles.

M. Momma. Ms. Momma was supported in her graduate studies with the assistance of this grant. She completed her dissertation in October 1990 and was awarded the degree of Ph.D. by Princeton University. He dissertation is entitled: "Estimation of Lifetime Models Using Intercepted Sampling Methods."

J. Gao. Mr. Gao was supported under this grant for two years of his graduate work at Princeton University. Mr. Gao completed his Princeton Ph.D in September, 1992, and his dissertation "Analysis of Two Heuristic Methods for the Euclidean Traveling Salesman Problem" has been communicated to the ARO Library. The final work on the thesis was complete under ARO Grant DAAL03-92-G-0110 and will be discussed more fully in the final report of that grant.

V. Related Work Currently Supported by ARO

This work reported on here has been developed further by the author under the support of Grant DAAL03-92-G-0110 of the same title. There was a delay in providing this final report that was occasioned by the transfer of the grant from Princeton University, and by the fact that the principal investigator was seriously injured in the Fall of 1991. The final report of Grant DAAL03-92-G-0110 has also been completed, and readers who are interested in the work reported here might also wish to review that second final report. There were also three modifications to DAAL03-89-G-0092, so the date of the present report is not as out-of-line as the original grant period would suggest.