DTIC FILE COPY

UNCLASSIFIED MASTER COPY - FOR REPRODUCTION PURPOSES

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

AD-A201 459

1b. RESTRICTIVE MARKINGS	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
3. DISTRIBUTION / AVAILABILITY OF REPORT	Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	
5. MONITORING ORGANIZATION REPORT NUMBER(S)	ARO 22565.8-MA
6a. NAME OF PERFORMING ORGANIZATION	University of Michigan
6b. OFFICE SYMBOL (If applicable)	
7a. NAME OF MONITORING ORGANIZATION	U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code)	Department of Statistics Ann Arbor, Michigan 48109-1027
7b. ADDRESS (City, State, and ZIP Code)	P. O. Box 12211 Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING / SPONSORING ORGANIZATION	U. S. Army Research Office
8b. OFFICE SYMBOL (If applicable)	
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	DAAG29-85-K-0008
8c. ADDRESS (City, State, and ZIP Code)	P. O. Box 12211 Research Triangle Park, NC 27709-2211

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Estimation in Large Samples

12. PERSONAL AUTHOR(S)

Michael Woodroofe

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 1/1/85 TO 6/30/88	August 1988	22

16. SUPPLEMENTARY NOTATION 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.

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Sampling, Estimation, Truncated Data, Asymptotic Expansions, Confidence Intervals, Optimal Sequential Designs

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Three problem areas were identified and studied: nonparametric estimation with truncated data; asymptotic expansions for randomly stopped sums (with a view towards finding corrected confidence intervals in sequential problems); and asymptotic local minimaxity in sequential estimation. In addition, some effort was devoted to understanding optimal sequential designs. The final report contains complete results of the research.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT

☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION

Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

ARO 22565.8

FINAL REPORT

ESTIMATION IN LARGE SAMPLES

DAAG29-85-K-0008

MICHAEL WOODROOFE: PROJECT DIRECTOR

CONTENTS

A-1

PROBLEMS STATEMENT

The proposal identified three problem areas: nonparametric estimation with truncated data; asymptotic expansions for randomly stopped sums (with a view towards finding corrected confidence intervals in sequential problems); and asymptotic local minimaxity in sequential estimation. In addition, some effort was devoted to understanding optimal sequential designs.

SUMMARY OF IMPORTANT FINDINGS

The problem of estimation with truncated data was studied in the thesis of Djamel Bellout. He considered the problem of searching for hidden objects of sizes $X_1,\ldots,X_N$, say. Let T_i be the time in which the i^{th} object would be found if the search were continued forever. Bellout supposed that the pairs $(X_1,T_1),\ldots,(X_N,T_N)$ are i.i.d., as (X,T) say, where X has a distribution function F, the conditional distribution of T given $X = x$ is exponential with failure H(x), and H is assumed to be non-decreasing. Here F, H, and N may all be unknown. If the process is observed for t time units, where $t > 0$, then the available data consist of those pairs (X_i,T_i) for which $T_i \le t$, a biased sample. Bellout was able to find nonparametric maximum likelihood estimates of the parameters F, H, and N and show their consistency. In addition, he studied an optimal stopping problem in which one receives a reward, depending on the sizes of the objects found. He first derived an optimal policy for the case of known F, H, and N, and then studied the adaptive policy in which estimators are substituted for the unknown parameters. Two papers have been submitted from this thesis, items 4 and 7 below.

Substantial progress was made on the problem of finding asymptotic expansions for the distributions of randomly stopped sums. The very weak expansions, described in the proposal, turn out to be a general technique for determining sequential confidence intervals. Their use for setting confidence intervals with fixed proportional accuracy for the parameter of an exponential family was studied in Refs 2 and 3 (below) and in the thesis of A. Meslem. The latter contains a complete asymptotic solution to the problem and extensive numerical calculations which indicate that the asymptotic formulas are nearly valid for moderate (expected) sample sizes. Current numerical work (still in progress) indicates that very weak expansions may be quite useful for setting confidence intervals after sequential testing.

Asymptotic local minimaxity in sequential estimation was studied extensively in the thesis of Mohamed Tahir. Early work by Woodroofe (1985 *Ann. Statist.*, 676-688) provided examples in which one stopping time, essentially a myopic procedure, was asymptotically second order optimal with respect to a large class of priors. This may be contrasted to the estimation component of the problem in which second order optimality requires knowledge of the prior and, thus, indicates that the sensitivity of an optimal design to the prior is less than the sensitivity of an estimator. Tahir was able to confirm this by extending to early result from the context of isolated examples to a much more general one.

During the period of the grant, it was realized that some of the mathematical techniques could be used in more general sequential design problems, problems in which there is a choice of experiment at each stage as well as optional stopping. Problems of this nature were studied in the theses of Kamal Rekab and Taoufik Zoubeidi. While limited to special cases, the results of these theses are encouraging. They indicate that a natural analogue of the myopic procedure (for stopping problems) may enjoy the same asymptotic optimality properties in more general design problems.

PERSONNEL SUPPORTED BY THE GRANT

Michael Woodroofe; project director and principle investigator.

Martha Aliaga; graduate student.

Djamel Bellout; graduate student.

Ki-Heon Choi; graduate student.

A. Meslem; graduate student.

Kamal Rekab; graduate student.

Taoufik Zoubeidi; graduate student.

PUBLICATIONS

The following is a list of all papers submitted and accepted, which were prepared with the support of the Grant. Abstracts of these papers may be found in Appendix 1.

1. Woodroofe, M. (1987). Confidence intervals with fixed proportional accuracy. *J. Statist. Plan. Inf., vol. 15, 131-146.*

2. Meslem, A. (1988). Asymptotic expansions for fixed width confidence intervals. *J. Statist. Plan. Inf., vol. 17, 51-66.* By A. Meslem.

3. Woodroofe, M. (1988). Fixed proportional accuracy in three stages. *Proceedings of the Fourth Purdue Symposium on Statistical Decision Theory and Related Topics, vol. 2*, 209-222.

4. Bellout, D. (1988). Order restricted estimation of distributions with censored data. To appear in the *J. Statist. Plan. Inf.*

5. Tahir, M. (1988). Asymptotically optimal Bayesian sequential point estimation with censored data. Tentatively accepted by *Sequential Analysis*.

6. Woodroofe, M. (1988). Corrected confidence intervals for adaptively designed experiments. Submitted to the *American Journal of Mathematical and Management Sciences* (Special issue in honor of Herb Robbins).

7. Bellout, D. (1988). An adaptive rule based on distribution estimates from truncated data. Tentatively accepted by *Sequential Analysis*.

8. Zoubeidi, T. (1988). Asymptotic efficiency of a sequential allocation rule. Tentatively accepted by *Sequential Analysis*.

DOCTORAL DISSERTATIONS

The following students wrote their theses with the partial support of the Grant. All have received a Ph.D. degree. Abstracts may be found in Appendix 2.

Martha Aliaga (July, 1986). *A problem in sequential analysis*.

Djamel Bellout (August, 1987). *Order restricted estimation of distributions with censored data and applications to a Stopping Problem*.

Choi, Ki-Heon. *An Adaptive Sequential Probability Ratio Test for Autoregressive Process.*

Meslem, A. (July, 1987). *Asymptotic expansions for confidence intervals with fixed proportional accuracy.*

Rekab, Kamal (August, 1988). *Asymptotic efficiency in sequential designs for estimation.*

Tahir, Mohamed (August, 1987). *Asymptotically optimal Bayes and minimax sequential point estimation.*

Zoubeidi, Taoufik (July, 1988). *Second order optimal sequential tests for clinical trials of two treatments.*

APPENDIX 1: ABSTRACTS OF PUBLICATIONS

item # 1

Journal of Statistical Planning and Inference 15 (1987) 131-146
North-Holland

CONFIDENCE INTERVALS WITH FIXED PROPORTIONAL ACCURACY*

Michael WOODROOFE

Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

Received 17 July 1985; revised version received 2 December 1985
Recommended by M.L. Puri

Abstract: The problem of setting a confidence interval with fixed proportional accuracy for the mean of a normal distribution is considered. A sequential procedure is proposed; and asymptotic expansions for its average coverage probabilities and its expected sample size are obtained. The procedure is shown to have asymptotically minimal expected sample size, subject to a constraint on the error probabilities.

AMS Subject Classification: 62L12.

Key words and phrases: Stopping times; Average confidence levels; Posterior distributions; Asymptotic expansions; Sequential decision problem.

1. Introduction

The possibility of using sequential methods to set a fixed width confidence interval for the mean of a distribution with an unknown variance has attracted substantial interest in theoretical statistics. This interest was stimulated by Stein's (1945) two-stage procedure, the fully sequential procedures of Anscombe (1953) and Chow and Robbins (1965) for the normal and non-parametric cases, and Hall's (1981) three stage procedure. The recent paper by Finster (1985) describes multiparameter extensions and may be consulted for further references.

Much of this research exploits the independence of the sample mean and variance in the normal case. The independence provides a simple relation between the coverage probability of a sequential procedure and the first two moments of the stopping time; and this relation permits a second order asymptotic expansion for the coverage probability as the width of the interval shrinks to zero. See, for example, Anscombe (1953), Hall (1981), and Woodroofe (1982, Section 10.2). The second order terms in these expansions are especially interesting, since the effect of optional stopping appears there.

* Research supported by the U.S. Army under DAAG29-85-K-0008.

ASYMPTOTIC EXPANSIONS FOR FIXED WIDTH CONFIDENCE INTERVAL

By

Abd--El-Hakim Meslem*

April 1986

The University of Michigan

Abstract

A sequential procedure for setting a fixed width confidence interval for the natural parameter of an exponential family is proposed and studied. Average confidence levels are considered instead of the classical confidence curves. The procedure is shown to perform very well in a simulation study.

Key words and phrases: stopping time: average coverage probability: Baysian model; excess over the boundary: asymptotic expansions.

AMS (1980) classification: 62L12

* Research supported by the U.S. Army under DAAG 29-85-K-0008

FIXED PROPORTIONAL ACCURACY IN THREE STAGES

MICHAEL WOODROOFE

Department of Statistics
The University of Michigan
Ann Arbor, Michigan 48109, U.S.A.

1. INTRODUCTION

Let $x_1, x_2, \ldots$ denote independent random variables which are normally distributed with an unknown mean $\theta \in R = (-\infty, \infty)$ and unit variance; let Q denote a positive, twice continuously differentiable function on R. A sequential confidence interval for θ with width proportional to $\sqrt{Q(\theta)}$ is sought; that is, given $0 < \gamma < 1$ and $h > 0$, a stopping time $t = t(x_1, x_2, \ldots)$ and an estimator $\theta_t = \theta_t(x_1, \ldots, x_t)$ are sought for which

$$P_\theta\{|\,\theta_t - \theta\,| < h\sqrt{Q(\theta)}\} \approx \gamma \tag{1}$$

for all θ. Small values of h are of special interest here, and $\approx$ is interpreted to mean equality up to $o(h^2)$. Woodroofe (1986a) considered this problem and established (1) in a very weak sense for a fully sequential procedure, one which monitors the data continuously. The main result of this paper is to show that (1) holds (in a conventional sense) for a three stage procedure, one which requires monitoring the data only three times. For simplicity, θ is estimated by the sample mean,

$$\theta_n = \bar{X}_n = (X_1 + \ldots + X_n)/n, \; n = 1, 2, \ldots, \tag{2}$$

although more general estimators are amenable to similar analyses, as in Woodroofe (1986a). Thus, the problem is to find a stopping time t for which (1) holds with θ_t as in (2).

To get a rough idea of the sample size required, let Φ denote the standard normal distribution function. If a sample of fixed (non-random) size n were taken, then the left side of (1) would be $2\Phi(h\sqrt{[nQ(\theta)]}\} - 1$, which exceeds $\gamma \iff$

$$n > N := c^2/h^2Q(\theta), \tag{3}$$

where $c = \Phi^{-1}[(1+\gamma)/2]$, and $:=$ indicates a definition. Of course, N depends on θ, which is unknown.

To describe the sequential procedure, let q denote a twice continuously differentiable function on R; let

$$Q_n = \max\{1/n, \; \min\{n, Q(\bar{X}_n)\}\}$$

item # 4

ORDER RESTRICTED ESTIMATION OF DISTRIBUTIONS WITH CENSORED DATA

by

DJAMEL BELLOUT*

September 1986

The University of Michigan

Abstract

Consider a finite population of hidden objects, and search for these objects for one unit of time. Suppose that both the size and the discovery time of the objects have unknown distributions, and that the conditional distribution of time given size is exponential with, as parameter, an unknown non-negative and non-decreasing function of the size. Order restricted M.L.E.'s are derived for this function, other parameters are estimated, and the consistency of the estimates is shown.

Key words and phrases: nonparametric, maximum likelihood estimation, order restriction, isotonic regression, least concave majorant, consistency.

AMS 1980 subject classifications: primary 62G05; secondary 62G30.

*Research supported by the U.S. Army under DAAG 29-85-K-0008.

i t, n 5

ASYMPTOTICALLY OPTIMAL BAYESIAN SEQUENTIAL POINT ESTIMATION WITH CENSORED DATA

Mohamed Tahir

Department of Statistics
The University of Michigan
Ann Arbor, MI 48109

Key Words and Phrases: posterior expected loss; Bayes estimator; Markov processes; infinitesimal operator; Dynkin's identity; asymptotic expansions.

ABSTRACT

A stopping time for estimating the failure rate of an exponential distribution with a smooth loss function and a gamma prior is constructed using censored data. The optimality of this stopping time is established by using Dynkin's identity for Markov processes. Although the results obtained are asymptotic in one sense, the optimality is exact.

1. INTRODUCTION

Let $X_1, \ldots, X_N$ be conditionally independent and exponentially distributed random variables with common failure rate ω, given that $W = \omega$, where W is a random variable having a gamma distribution with parameters a_0 and b_0. Further, let

$$M(t) = \sum_{i=1}^{N} I_{[0,\, t]}(X_i)$$

denote the number of observations which have been made by time $t \geq 0$, where I denotes the indicator function. Also, let F_t be the sigma-algebra generated by $M(s)$, $0 \leq s \leq t$ and $Y_1, \ldots, Y_{M(t)}$, where $Y_1, \ldots, Y_N$ are the order statistics of $X_1, \ldots, X_N$. It follows that the likelihood function at time t is proportional to

$$l(\omega \mid t) \propto \omega^{M(t)} \exp\{-\omega \sum_{i=1}^{M(t)} Y_i - \omega\, [N - M(t)]t\}$$

for $\omega > 0$. Thus, the posterior distribution of W, given F_t, is a gamma distribution with new parameters

Item 6

CORRECTED CONFIDENCE LEVELS

FOR

ADAPTIVELY DESIGNED EXPERIMENTS

By

Michael Woodroofe*
The University of Michigan

ABSTRACT

Consider a non-linear regression model $y_k = g(x_k;\omega) + e_k$, $k = 1,2,\ldots,$ in which the design variables, $x_k = x_k(y_1,\ldots,y_{k-1})$, may be functions of the previous responses. A very weak asymptotic expansion for the distribution of the maximum likelihood estimator is presented, from which the primary effect of the adaptive nature of the design may be seen.

Key words and phrases: non-linear regression; stable designs; martingale central limit theorem; very weak expansions.

* Research supported by the U.S. Army Research Office under DAAG 29-85-0008.

AN ADAPTIVE RULE BASED ON DISTRIBUTION ESTIMATES FROM TRUNCATED DATA

Djamel Bellout*

Department of Statistics/C & IS
The George Washington University
Washington, D. C. 20052

Key words and phrases: payoff; Markov process; characteristic operator; stopping time; adaptive rule.

ABSTRACT

Consider a finite population, of large but unknown size, of hidden objects. Consider searching for these objects for a period of time, at a certain cost, and receiving a reward depending on the sizes of the objects found. Suppose that the size and discovery time of the objects both have unknown distributions, but the conditional distribution of time given size is exponential with an unknown non-negative and non-decreasing function of the size as failure rate. The goal is to find an optimal way to stop the discovery process. First, assuming that the above parameters are known, an optimal stopping time is derived and its asymptotic properties are studied. Then, an adaptive rule, based on order restricted estimates of the distributions from truncated data, is presented. This adaptive rule is shown to perform nearly as well as the optimal stopping time, for large population size.

1. INTRODUCTION

Consider a finite population, of large but unknown size N, of hidden objects. Consider searching for these objects for a period $t > 0$ of time, at a certain cost pro-

ASYMPTOTIC EFFICIENCY OF A SEQUENTIAL ALLOCATION RULE

Taoufik Zoubeidi

Department of Statistics
University of Michigan
Ann Arbor, MI 48105

Key Words and Phrases: sequential clinical trials; hypothesis testing; optimal stopping; dynamic programming.

ABSTRACT

In a clinical trial of two treatments, one goal of the experimenter is to design an experiment such that the number of patients assigned to the inferior treatment is minimized. A Bayesian formulation for this problem is studied. Treatment outcomes are assumed to be independent and normally distributed, with conjugate priors. The analysis is conducted in a large sample limit where sampling costs are scaled to zero. First we show there is little harm restricting attention to procedures that stop according to Kiefer and Sacks stopping rule, there are procedures in this class with risks that exceed the Bayes risk by the cost of a fixed number of observations. Using this we then develop fully efficient procedures with risks asymptotic to the Bayes risk in our large sample limit.

APPENDIX 2: ABSTRACTS OF DISSERTATIONS

A PROBLEM IN SEQUENTIAL ANALYSIS

by

Martha Beatriz Bilotti-Aliaga

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
1986

Doctoral Committee:

Professor Michael B. Woodroofe, Chairman
Assistant Professor Robert Keener
Professor Robb Muirhead
Professor Maxwell Reade
Professor Edward Rothman

ABSTRACT

ORDER RESTRICTED ESTIMATION OF DISTRIBUTIONS WITH CENSORED DATA AND APPLICATION TO A STOPPING PROBLEM

by
Djamel Bellout

Chairperson: Michael B. Woodroofe

Consider a finite population of hidden objects with sizes $X_1, X_2, \ldots, X_N$. Let T_i denote the time at which the i–th object would be found in an infinite search. Let F be the unknown distribution of X; G^* the conditional distribution of T given X. Suppose that G^* is of the form $G^*(t|x) = 1 - \exp\{-tH(x)\}$, where H is an unknown, non–negative, and non–decreasing function. Finally, suppose that only the pairs (X_i, T_i) for which $T_i \leq t$, $t > 0$ are observed. An optimal stopping time is constructed for the stopping problem with the payoff $Z_u(t) = \sum_{i=1}^{N} u(X_i) I\{T_i \leq t\} - ct$, where u is a non–decreasing function and c is the cost of sampling per unit time, when F, H and N are known. Maximum likelihood estimates of F, H and N are obtained, using techniques of isotonic regression. These are shown to be strongly consistent, and some remarks are made about their asymptotic distribution. An adaptive stopping time in which F, H and N are replaced by estimates is proposed. The adaptive rule is shown to perform as well as the optimal stopping rule, asymptotically.

ABSTRACT

AN ADAPTIVE SEQUENTIAL PROBABILITY RATIO TEST FOR AUTOREGRESSIVE PROCESS

by
Ki-Heon Choi

Chairperson: Michael B. Woodroofe

Consider the first order autoregressive model

$$X_n = \theta + \rho X_{n-1} + \epsilon_n, \qquad n \in \mathcal{Z} \quad .$$

where ϵ_n, $n \in \mathcal{Z}$, are i.i.d. normally distributed random variables with mean 0 and variance σ^2. Here σ^2, θ, and ρ are parameters. we consider the problem of sequentially testing hypotheses about a parameter θ in the presence of the nuisance parameter ρ .

The sequential testing problem leads directly to the study of stopping times of the form

$$t = t_a = \inf\{ n \geq 1 : S_n > a \}$$

where S_n is a function of $X_1, \cdots, X_n$ and $a > 0$, and the problem of evaluating of conditional boundary crossing probabilities of the form

$$\Pr(t_a \geq n \mid S_n = a + r)$$

for a , $r > 0$ and $n = 1, 2, \cdots$. The approach is to compute the conditional probability of crossing the boundary a , looking backward along the sequence $S_{n-1}, \cdots . S_{n-k}$, by obtaining local limit theorems for the distribution of $S_n, S_{n-1}, \cdots, S_{n-k}$.

ABSTRACT

ASYMPTOTICALLY OPTIMAL BAYESIAN AND MINIMAX SEQUENTIAL POINT ESTIMATION

by
Mohamed Tahir

Chairperson: Michael B. Woodroofe

Four related problems are considered in this thesis. The first problem consists of determining an optimal stopping time for estimating the failure rate of an exponential distribution with a general, smooth loss function and a gamma prior, using censored data. The construction of such a stopping time involves obtaining asymptotic expansions for the posterior expected loss, computing the infinitesimal operator of some resulting Markov process, and using Dynkin's identity. The second problem is to obtain an asymptotic lower bound for the minimax regret of a sequential estimation procedure, in the case of a one-parameter exponential family, for a general class of estimators of the population mean and with a general, smooth loss function. The program is to determine the limit of the Bayes regret and then maximize with respect to the prior distribution. The third problem is to develop a uniform version of the non-linear renewal theorem. Finally, the fourth problem is to find a stopping time which attains the asymptotic lower bound of the second problem. The procedure used to yield such a stopping time requires the results developed in the third problem.

ABSTRACT

ASYMPTOTIC EFFICIENCY IN SEQUENTIAL DESIGNS FOR ESTIMATION

by
Kamel Rekab

Chairperson: Michael B. Woodroofe

Properties of a sequential design for nonlinear regression and related problems are investigated. The main results determine an asymptotic lower bound for the Bayes risk. and a fully sequential procedure as well as a two stage procedure that are first order efficient.

ABSTRACT

SEQUENTIAL TESTS FOR CLINICAL TRIALS OF TWO TREATMENTS WITH BOUNDED DEFICIENCY

by
Taoufik Zoubeidi

co-Chairpersons: Robert W. Keener and Michael B. Woodroofe

In a clinical trial of two treatments, one goal of the experimenter is to design an experiment such that the number of patients assigned to the inferior treatment is minimized. A Bayesian formulation for this problem is studied. Treatment outcomes are assumed to be independent and normally distributed, with conjugate priors. The analysis is conducted in a large sample limit where sampling costs are scaled to zero. First we show there is little harm restricting attention to procedures that stop according to Kiefer and Sacks stopping rule. There are procedures in this class with risks that exceed the minimum Bayes risk by the cost of a fixed number of observations. Using this we then develop procedures with bounded deficiency. A Monte Carlo study shows that these procedures perform better than the pairwise procedure.