AD-A268 895

REPORT DOCUMENTATION PAGE

OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	FINAL/01 MAY 89 TO 31 DEC 92

4. TITLE AND SUBTITLE

SOFTWARE RELIABILITY: ESTIMATION & PREDICTION (U)

5. FUNDING NUMBERS

AEOSR-TR- 93 0657

2304/A5

6. AUTHOR(S)

Dr Gordon Kaufman

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

Massachusetts Institute of Technology
MIT School of Management
Newark, DE 19716

8. PERFORMING ORGANIZATION REPORT NUMBER

DTIC
ELECTE
SEP 0 2 1993
S B D

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

AFOSR/NM
110 DUNCAN AVE, SUTE B115
BOLLING AFB DC 20332-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFOSR-89-0371

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED	UL

13. ABSTRACT (Maximum 200 words)

The principal purpose of this research was to provide new tools for measuring probabilities of failure free operation of software systems and to develop methods for predictions of software reliability. Links are established between stochastic models of fault occurrence suggested by Scholtz (1985) and Miller (1986) and an important class of finite population sampling models called "successive sampling" in the sample survey literature. Successive sampling consists of sampling a finite population of objects, each with an assigned magnitude, proportional to magnitude and without replacement. Recognition of linkages between Schoz and Miller's "Exponential Order Statistics Models" and successive sampling allows application of an emerging body of research on methods of estimation for successive sampling models to software reliability estimation.

93-20499

93 01 029

14. SUBJECT TERMS	15. NUMBER OF PAGES
	18
	16. PRICE CODE

17. SECURITY CLASSIFICATION	18. SECURITY CLASSIFICATION	19. SECURITY CLASSIFICATION	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR(SAME AS REPORT)

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

(11954)
P.1
Accept:

AFOSR Contract #AFOSR-890371 FINAL REPORT

SOFTWARE RELIABILITY: ESTIMATION AND PREDICTION

Gordon M. Kaufman

The principal purposes of this research was to provide new tools for measuring probiabilities of failure-free operation of software systems and to develop methods for predictions of software reliablity.

Links are established between stochastic models of fault occurrence suggested by Scholz (1985) and Miller (1986) and an important class of finite population sampling models called "successive sampling" in the sample survey literature. Successive sampling consists of sampling a finite population of objects, each with an assigned magnitude, proportional to magnitude and without replacement. Recognition of linkages between Schoz and Miller's "Exponential Order Statistics models" and successive sampling allows application of an emerging body of research on methods of estimation for successive sampling models to software reliability estimation.

Two papers are devoted to extensions of the theory of exponential order statistics models and to presentation of methods of estimation based on a data record of times to failures. A novel feature is the development of methods of estimation that maintains the distinction between types of software failures (logic, coding, interface, etc.) In particular, given a data record of both the type of each observed software failure and the time at which it occurred, the question of how to estimate the number of each fault type remaining in the system and the time on test needed to discover some fraction of these remaining faults is addressed. Estimation methods studies are maximum likelihood, conditional maximum likelihood and unbiased estimation:

"Software Reliability Modeling and Exponential Order Statistics" MIT Sloan School Working Paper 3114-90MS, January 1990 (with G. Andreatta) 45 pp.

"Successive Sampling and Software Reliability" Sloan School Working Paper 3316, July 1991 (In review with IEEE Transactions on Software Engineering) 29 pp.

Profile maximum likelihood methods for estimating remaining faults by type in NASA/Goddard SEL software test data were presented at a TIMS conference in November 1992. This paper is in progress along with a paper on a Bayesian treatment of successive sampling inference, entitled "Bayesian Successive Sampling Inference". An invited presentation on the latter topic was presented at the Latin-American-U.S. Workshop on Bayesian Statistics and Econometrics in Caracas, Venezuela, December 9-14, 1992.

At termination of this contract, development of efficient numerical schemes for solution of non-linear efficient score functions for profile maximum likelihood and work on Bayesian alternatives for estimation in light of observation of NASA/Goddard type data is under way.

TIMS NOV. 1992

SUCCESSIVE SAMPLING (SS) AND SOFTWARE RELIABILITY

OBJECTIVES:

- ❑ SHOW THAT EXPONENTIAL ORDER STATISTICS MODELS (EOS) = SS
- ❑ HOW SS ESTIMATION METHODS CAN APPLY TO SOFTWARE RELIABILITY TO:
- ❑ ESTIMATE RETURNS TO TESTING EFFORT

GIVEN OBSERVED TEST HISTORY:

(A) HOW MANY FAULTS OF WHAT TYPE REMAIN?

(B) HOW MUCH ADDED TIME ON TEST IS NEEDED TO UNCOVER m MORE FAULTS?

(C) IF WE TEST FOR T MORE UNITS OF TIME, HOW MANY FAULTS OF WHAT TYPES WILL BE OBSERVED?

EXAMPLE: ERBS PROJECT (NASA - GOODDARD - SEL)

SEL-GODDARD MODEL

I. FAULT TYPES AREA DISTINGUISHABLE

II. RELIABILITY GROWTH CAPTURED BY SS

II. THE # OF FAULTS OF EACH TYPE ARE SUPER-POPULATION GENERATED

IV. THE SUPER-POPULATION IS NON-PARAMETRIC

SS ⊕ SUPERPOPULATION PROCESS

I) GIVEN $A_N = \{a_1, ..., a_N\}$

$$\text{PROB}\{s_n | A_N\} = \prod_{j=1}^{n} \frac{a_{i_j}}{a_{i_j} + \cdots + a_{i_N}}, \quad n \leq N$$

(II) $a_1, .., a_N$ ARE VALUES OF N MID RVS $A_1, ..., A_N$ WITH COMMON CDF $F(\cdot|\underline{\theta})$ CONCENTRATED ON $(0, \infty)$

⇓

$$\text{PROB}\{A_j = a_k\} = \theta_k, \quad k=1,2,..., k$$

$$\sum_{k=1}^{K} \theta_k = 1, \quad \theta_k \geq 0.$$

ANALYSIS

(1) A NON-PARAMETERIC PROFILE ML ESTIMATE OF PROPORTIONS OF EACH FAULT TYPE

(2) BOUNDS ON PARAMETER ESTIMATES

(3) PROFILE MLE FOR NUMBER OF FAULTS BY TYPE

HOW MANY FAULTS REMAIN?

PROFILE MLE REMAINING FAULTS*

	f =					
	1.0	.9	.8	.7	.6	.5
COMPUTE	0	2	6	9	15	23
DATAVAL	0	3	6	11	16	23
INIT	0	1	3	5	9	15
INTERE	0	1	3	4	6	8
INTERI	0	4	1	2	2	3
LOGIC	0	4	10	18	30	50
TOTAL REMAINING	0	15	29	50	78	122
$N - n = n\left(\frac{1-f}{f}\right)$	0	13	29	49	76	114

*FOR K = 2 ONLY!

COMPLICATED ANALYSIS YIELDS SIMPLE BENCHARKS

$$f\,\bar{r}_k \le \hat{\theta}_k \le 1 - f + f\,\bar{r}_k$$

$$f = \frac{n}{N} \qquad \bar{r}_k = \frac{r_k}{n}$$

FAULT TYPE	RANGE OF $\hat{\theta}_k$			
	$f = 1.0$	$f = .80$	$f = .50$	$f = .20$
COMPUTE	[.1842, .1842]	[.1474, .3474]	[.0921, .5921]	[.0368, .8368]
DATAVAL	[.2456, .2456]	[.1965, .3965]	[.1228, .6228]	[.0491, .8491]
INIT	[.0789, .0789]	[.0631, .2631]	[.0395, .5395]	[.0158, .8158]
INTERE	[.1053, .1053]	[.0842, .2842]	[.0527, .5527]	[.0211, .8211]
INTERI	[.1316, .1316]	[.1052, .3052]	[.0658, .5658]	[.0263, .8263]
LOGIC	[.2544, .2544]	[.2035, .4035]	[.1272, .6272]	[.0509, .8509]

↑

SUMS TO 1.000

SOFTWARE RELIABILITY MODELING AND EXPONENTIAL ORDER STATISTICS

by

Giovanni Andreatta and Gordon M. Kaufman

MIT Sloan School Working Paper 3114-90MS
January 1990

SOFTWARE RELIABILITY MODELING AND EXPONENTIAL ORDER STATISTICS

by

Giovanni Andreatta and Gordon M. Kaufman*

ABSTRACT: Properties of software failure times modelled as realizations of order statistics generated by independent but non-identically distributed exponential random variables are developed. Edgeworth and saddle point approximations to central order statistic densities so generated are developed using an exact integral representation of these densities. A comparison of Edgeworth and saddle point approximation with exact densities for two different population types is given. The accuracy of the saddle point approximation, even for very small population sizes ($N = 6$) and small samples ($n = 2$) is excellent.

The same technique is used to provide an exact integral representation of the probability that a particular fault appears in a sample of a given size. Some numerical comparisons of Rosén's (1972) approximation of inclusion probabilities with exact values are provided. His simple approximation appears to give excellent results as well.

The intimate connection between successive sampling theory and EOS models for software reliability is documented.

KEY WORDS: SOFTWARE RELIABILITY, SUCCESSIVE SAMPLING, EDGEWORTH APPROXIMATION, SADDLE POINT APPROXIMATIONS, INCLUSION PROBABILITY, ORDER STATISTICS

* Supported by AFOSR Contract #AFOSR-89-0371

1. INTRODUCTION

Goel (1985) has defined software reliability as the probability that during a prespecified testing or operational time interval, software faults do not cause a program to fail:

> "Let F be a class of faults, defined arbitrarily, and T be measure of relevant time, the units of which are dictated by the application at hand. Then the reliability of the software with respect to the class of faults F and with respect to the metric T, is the probability that no fault of the class occurs during the execution of the program for a prespecified period of relevant time."

Several classes of models have been proposed to capture this definition of reliability; among the most prominent are models built on the assumptions that waiting times between software failures are exponentially distributed and in addition are, conditional on knowledge of the appropriate parameter set, mutually independent. Such models have been called Exponential Order Statistics (EOS) models by Miller (1986) in his investigation of similarities of and differences between models based on the aforementioned assumptions. Littlewood (1981) was perhaps the first to challenge the assumption adopted by many authors that each fault "...contributes the same amount to the overall failure rate..." She posits a model in which (a) each fault possesses a parameter (occurrence rate) individual to that fault and (b) the collection of fault parameters is generated by a superpopulation process. This approach has the decisive advantage of avoiding some analytical and computational complexities that arise when assumption (b) is dropped. It is empirical Bayes in spirit and so is in formal correspondence with the Bayesian approach to reliability modeling adopted by Singpurwalla and his co-authors (Langberg and Singpurwalla (1985)) for example. However, Miller argues that Littlewood's model minus the assumption (b), a model that he calls a deterministic EOS model, "...has a certain physical motivation: the individual failure rates are physical quantities in the sense that they can be estimated to any desired degree of accuracy. The IDOS [empirical Bayes] and NHPP [non-homogeneous Poisson process] models are attractive because of mathematical tractibility and successful application experience; however, they are more difficult to motivate and verify in a physical sense." (Miller (1986), p. 12). In sum, some researchers view the EOS model as a first principles model that captures the physics of fault occurrence more accurately than the alternatives explored in the literature. This led Miller (1986) and Scholz (1986) to explore properties of order statistics generated by mutually independent but non-identically distributed random

variables – the analytical concommitant of the EOS model.

The connection of this line of research with a sampling scheme well known to sample survey statisticians – successive sampling or sampling proportional to magnitude and without replacement from a finite population of magnitudes – has passed unnoticed until now. One of the purposes of this paper is to establish the nature of this connection. The problem of making inferences about unobserved finite population parameters of the EOS model based on observation of waiting times between failures and possibly the magnitude of observed faults is a dual of the problem of inference based on observation of fault magnitudes alone. The later problem has been investigated in detail by several authors (Andreatta and Kaufman (1986); Gordon (1989); Wang and Nair (1986); Bickel, Nair, and Wang (1989)). Other features of the link between software reliability models and successive sampling appear in a companion paper (Kaufman (1989b)).

Another purpose is to provide tools for the computation of the distribution of central order statistics for the EOS model and for the probability that a fault possessing a prespecified magnitude will be included in a sample of faults of a given size. Both play an important role in theories of inference for EOS models. The distribution of the waiting time to occurrence of the nth fault is an analytical benchmark for understanding properties of the EOS model and for a theory of unbiased estimation of the empirical distribution of magnitudes of unobserved faults and of the number of faults remaining in the software system.

Gordon (1982) has shown that the distribution of permutations of the order in which successively sampled elements of a finite population are observed can be characterized in terms of exponential waiting times with expectations inversely proportional to magnitudes of the finite population elements. This leads naturally to a corollary interpretation of the probability that a particular element of the population will be included in a sample as the expectation of an exponential function of an order statistic generated by independent but non–identically distributed exponential random variables (rvs).

In Section 3 we present an exact integral representation of the marginal density of an order statistic so generated. The integrand is interpretable as a probability mixture of characteristic functions of sums of conditionally independent Bernoulli rvs, an interpretation that suggests a first approximation of the density, and the form that leading terms in Edgeworth and saddle–point approximations will take.

An Edgeworth type approximation is presented in Section 4. While this expansion could in principle be derived by first computing a saddle–point approximation and then using the idea of recentering a conjugate distribution as suggested by Daniels (1954), we have chosen to compute it directly.

SUCCESSIVE SAMPLING AND SOFTWARE RELIABILITY

by

Gordon M. Kaufman*

MIT Sloan School Working Paper 3316
July 1991
Revised October 1992

*Supported by AFOSR Contract #AFOSR-890371. I wish to thank Nancy Choi and Tom Wright for valuable programming assistance and Chris Kemerer for insightful comments.

SUCCESSIVE SAMPLING AND SOFTWARE RELIABILITY

by

Gordon M. Kaufman*

1. Introduction

A software system is tested and times between failures are observed. How many faults remain in the system? What is the waiting time to the next failure? To the occurrence of the next n failures? Conditional on the observed history of the test process, knowledge of properties of the time on test necessary to discover the next n faults is very useful for making test design decisions.

Times between failures models of software reliability are designed to answer such questions. Many versions of such models appear in the literature on software reliability and most such models rely on the assumption that the failure rate is proportional to the number of remaining faults or to some single valued function of the number of remaining faults. Goel (1985) observes that this is a reasonable assumption if the experimental design of the test assures equal probability of executing all portions of the code -- a design seldom achieved in practice. The character of testing usually varies with the test phase: requirements, unit, system or operational. The impact of such considerations have been recognized by some authors: Littlewood's criticism of the Jelinski-Moranda assumption that software failure rate at any point in time is directly proportional to the residual number of faults in the software is cited by Langberg and Singpurwalla (1985) in an excellent overview paper. Only recently have some researchers come to grips with the implications of replacing this assumption. In terms of counts of failures, it may be labelled an "equal bug size" postulate (Scholz (1986). Littlewood (1981) and Langberg and Singpurwalla (1985) do incorporate the assumption that different bugs may have different failure rates, but the empirical Bayes (superpopulation) approach adopted by Littlewood and the Bayesian approach adopted by Singpurwalla and Langberg "averages out" the effects of this assumption. According to Scholz "...it was not recognized by some proponents of reliability growth models that relaxing the equal bug size assumption also entails some complications concerning the independence and exponentiality [of waiting

* Supported by AFOSR Contract #AFOSR-89-0371. I wish to thank Nancy Choi and Tom Wright for valuable programming assistance and Chris Kemerer for insightful comments.

times between failures]". He and Miller (1986) are the first to investigate systematically (in the absence of a superpopulation process or of a Bayesian prior for failure rates) the implications of assuming that given an observational history, each of the remaining bugs in a software system may possess different probabilities of detection at a given point in time. In contrast to most times between failures models, for successive sampling - EOS models, times between failures are not independent. As Goel [1985] points out, independence would be acceptable if "...successive test cases were chosen randomly. However, testing especially functional testing, is not based on independent test cases, so that the test process is not likely to be random".

Scholz presents a multinomial model for software reliability that is identical to Rosén's characterization of successive sampling stopping times. (Rosén, 1972) The connection seems to have gone unnoticed. The "continuous" model based on independent, non-identically distributed exponential random variables suggested by Scholz as an approximation to multinomial waiting times is in fact in exact correspondence with a representation of successive sampling in terms of non-identically distributed but independent exponential order statistics. Scholz's approximation is in fact Ross's (1985) exponential order statistics model which Ross treats Bayesianly. Gordon (1983) was among the first to observe that successive sampling is representable in this fashion. Miller's study of such order statistics is focused on similarities and differences between types of models derivable from this particular paradigm.

Joe (1989) provides an asymptotic (large sample) maximum likelihood theory for parametric order statistics models and non-homogeneous Poisson models of fault occurrence that, when the parameter is of fixed dimension, yields asymptotic confidence intervals. He states that for the general exponential order statistics model, one cannot expect any estimate [of the conditional failure rate] to be good because the ratio of parameters to random variables is too big".

Successive sampling as described in the next section has been successfully used as a model for the evolution of magnitudes of oil and gas field discovery and has its roots in the sample survey literature. (Hájek (1981), for example.) In this application magnitudes of fields in order of discovery are observed and used to make predictions of the empirical frequencies of magnitudes of undiscovered fields. Logically tight theories of maximum likelihood, moment type and unbiased estimation for this class of problems have been developed by Bickel, Nair and Wang, (1992), Gordon, (1992) and Andreatta and Kaufman, (1986). The problem of estimation of software reliability based on observation of times between failures of a software system may be viewed as the dual to the problem of inference when only magnitudes of population elements are observed. The principal purpose of this

paper is to establish connections between these two disparate lines of research and to lay out possibilities for applying methods of estimation developed for successive sampling schemes to successive sampling as a model for software reliability. Our attention is restricted to successive sampling of elements of a finite population of software faults in a software system; that is,

(1) Individual faults may possess distinct failure rates that depend on covariates particular to the stage of testing and on other features of the software environment. For a given fault, that fault's failure rate as a function of such covariates is called the fault **magnitude**.

(2) Faults are sampled (a) without replacement and (b) proportional to magnitude.

Some recent studies of successive sampling schemes have assumed the existence of a superpopulation process generating finite population magnitudes. We shall not.

The accuracy of model structure as a depiction of the physics of software fault occurrence depends in part on the validity of choice of definition for the magnitude of a fault. Different definitions of magnitude may be required for different environments. Here we shall assume that the appropriate definition of the magnitude of a fault for the particular application considered has been resolved. It is NOT easy to resolve and considerable effort must be devoted to defining operationally meaningful definitions of fault magnitudes. An example will help to clarify the meaning of "fault magnitude". The Software Engineering laboratory at NASA-Goddard has gathered detailed data from six software projects. For some of these projects, failure data blocked by test phase is recorded in a form that displays failures by type as a function of cumulative hours on test. Six distinct failure types labelled "compute", "dataval", "init" "intere", "interi" and "logic" are distinguished in the ERBS project acceptance phase, for example. The number of failures of each type that occurred within each week of ten weeks of acceptance phase testing are recorded along with the weekly number of time on test hours expended by all programmers working on this test phase. In the context of the successive sampling model of fault occurrence (defined in the next section), each of these six distinct failure types is associated with a positive number; let a_i be the number associated with failure type $i = 1, \ldots, 6$. The **magnitude** a_i of fault type i may be interpreted as the reciprocal of the expectation of an exponential random variable belonging to each fault of type i. In turn, a_i may be made to be a function of one or more directly observed attributes that covary with the type of fault; e.g. each a_i may be a function of programmer time necessary to fix faults of type i. Empirical work by Basili and Perricone (1982) and Basili and Patniak (1986) provides an excellent starting point for study of how to define such covariates in an operationally meaningful way. But

this is a subject for a different paper.

Following a formal description of successive sampling properties of successive sampling schemes needed in the sequel, two distinct sampling (observational) schemes are examined in section three. The first is a scheme in which both the time from start of testing to time of occurrence and the magnitude of each fault in a sample of n faults are jointly observed. With this scheme we can order faults observed from first to last and assign a "waiting time" to each fault. In the second scheme magnitudes of faults in a sample of n faults are observed along with the waiting time to occurrence of the last fault in the sample; waiting times to occurrences of individual faults are not observed. The order in which faults occurred is then lost.

Section 4 is devoted to properties of unbiased estimators of unobserved finite population parameters for each of the two aftermentioned sampling schemes. The connection between maximum likelihood estimation (MLE) and unbiased estimation established by Bickel, Nair and Wang (1992) for a successive sampling scheme in which magnitudes alone are observed is developed for a scheme in which both waiting times to failures and magnitudes are observed. The results of a Monte Carlo study of properties of both types of estimators presented in Section 5.

Section 6 returns a principal interest of the software manager: conditional on observing the history of the process up to and including the m^{th} failure, what is the waiting time to the occurrence of the next $n - m$ failures? Successive sampling theory suggests a simple point estimator of this waiting time, dependent on the waiting time $z_{(m)}$ to occurrence of the first m faults and on the unordered set $\{y_1, \ldots, y_m\}$ of magnitudes of faults observed in $(0, z_{(m)})$. A Monte Carlo study of its behavior suggests that this class of estimators of returns to test effort measured in faults/unit time on test is worth further study.