FURTHER STUDIES IN ESTIMATION OF LIFE DISTRIBUTION CHARACTERISTICS FROM C. (U) SOUTH CAROLINA UNIV COLUMBIA DEPT OF STATISTICS W J PADGETT JUN 87 UNCLASSIFIED ARO-21245 26-MA MIPR-ARO-139-85 F/G 12/3 NL

W. J. Padgett

Department of Statistics
University of South Carolina
Columbia, SC 29208

U. S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

Approved for public release; distribution unlimited.

The main objectives of this research have been the development of smooth nonparametric estimators of quantile functions from right-censored data and the further study of smooth density estimators from censored observations. In particular, kernel-type quantile estimators and "generalized" product-limit quantile estimators have been obtained under censoring which give better estimates of percentiles of the lifetime distribution than the usual product-limit quantile estimator. Asymptotic properties of these quantile estimators have been developed, including asymptotic normality and consistency. Mean square convergence for...
the kernel estimator was studied. In addition, a data-based procedure for selecting the bandwidth has been investigated using the bootstrap, and approximate confidence intervals for the true quantile have been proposed using bootstrap estimates of the sampling distribution. Theoretical results on the optimal bandwidth selection for kernel density estimators under random censorship have also been obtained.

New results in several other problem areas were also developed. These included the study of linear empirical Bayes estimators, prediction intervals for the inverse Gaussian distribution, nonparametric hazard rate estimation under censoring, nonparametric inference for step-stress accelerated life tests under censoring, discrete failure models, simultaneous confidence intervals for pairwise differences of normal means, and optimal designs for comparing treatments with a control. In addition, reliability estimation when the cause of failure may be only partially known and problems giving rise to Gompertzian failure models were studied.
FURTHER STUDIES IN ESTIMATION OF LIFE DISTRIBUTION CHARACTERISTICS FROM CENSORED DATA: FINAL TECHNICAL REPORT (6-1-84 to 5-31-87)

W. J. Padgett, Principal Investigator

Grant Number MIPR ARO 139-85

Department of Statistics
University of South Carolina
Columbia, SC 29208

June 1, 1987
1. SUMMARY

This document reports the work performed and other research activities of the investigators during the funding period from June 1, 1984, to May 31, 1987, under grant number MIPR ARO 139-85.

The main objectives of this research have been the development of smooth nonparametric estimators of quantile functions from right-censored data and the further study of smooth density estimators from censored observations. In particular, kernel-type and "generalized" quantile estimators have been obtained under censoring which give better estimates of percentiles of the lifetime distribution than the usual product-limit quantile estimator. Asymptotic properties of these quantile estimators have been developed, including asymptotic normality and consistency. Also, the mean-square convergence and mean squared error of the kernel quantile estimators have been investigated. A data-based procedure for selecting the bandwidth has been studied using the bootstrap, and approximate confidence intervals for the true quantile have been proposed using bootstrap estimates of the sampling distribution. Theoretical results on the optimal bandwidth selection for kernel density estimators under random right censorship have also been obtained.

New results in several other problem areas were also developed. These included the study of linear empirical Bayes estimators, prediction intervals for the inverse Gaussian distribution, nonparametric hazard rate estimation under censoring, nonparametric inference for step-stress accelerated life tests under censoring, discrete failure models, reliability estimation when cause of failure is partially known, Gompertzian failure models, simultaneous confidence intervals for pairwise differences of normal means, and optimal designs for comparing treatments with a control. All of these results should prove
fruitful in the assessment of reliability and maintenance policies of various types of equipment.

2. RESEARCH OBJECTIVES

The research objectives of this project can be divided into the following categories:

A. Nonparametric estimation of quantiles under censoring;
B. Kernel density estimation from randomly right-censored data;
C. Nonparametric estimation from step-stress accelerated life tests;
D. Linear empirical Bayes estimation of mean time to failure; and
E. Miscellaneous problems in reliability.

The specific research problems that were considered during this project period in these areas will be briefly outlined below. The results obtained will be described in Section 3 of this report.

A. Nonparametric estimation of quantiles under censoring. One of the main objectives of this research was to develop smooth nonparametric estimates of the quantile function (or percentiles) from randomly right-censored data. In particular, the kernel-type quantile estimators were to be studied with respect to small-sample, as well as asymptotic, properties. Also, a modification of the kernel estimator which allowed the bandwidth to be a function of the censored data was to be studied. A different approach to smooth quantile estimation was taken in developing a "generalized" product-limit quantile estimator. This estimator is essentially an average of product-limit quantiles from all possible subsamples of a given size. These types of estimators smooth the product-limit quantile function 
\[ \hat{Q}_n(p) = \inf \{ t : \hat{F}_n(t) > p \} \]

where \( \hat{F}_n \) denotes the product-limit estimator of the lifetime distribution \( F_0 \), and are better than \( \hat{Q}_n(p) \) in the sense of smaller
mean squared errors in many situations. Thus, such estimators can provide more accurate estimates of percentiles, aiding in the assessment of equipment reliability.

B. Kernel density estimation from randomly right-censored data. Another objective of this project during the reporting period was to determine a method of "optimal" bandwidth selection for kernel-type density estimators under random right-censorship, with respect to some optimality criterion. The kernel density estimators of Blum and Susarla (Multivariate Analysis-V, 1980) and Földes, Rejtö and Winter (Periodica Mathematica Hungarica, 1980) were studied.

C. Nonparametric estimation from step-stress accelerated life tests. Usually in accelerated life tests, items are put on test at several accelerated stresses and observed until failure or until the test terminates. A more realistic method is to place items on test and allow the stress levels to change at preassigned times. This will result in more failures in general. The objective of the research in this area was to obtain a strongly consistent nonparametric estimator of the unknown life distribution at the normal use stress. A rescaling technique was utilized to study this problem.

D. Linear empirical Bayes estimation of mean time to failure. In estimating the mean lifetimes of n independent problems, it is hoped that the linear empirical Bayes estimators will converge to the optimal linear Bayes estimators with reasonable rapidity. This convergence was to be studied here. Also, another objective in this area was to identify a natural class of estimators between the linear Bayes estimators and the Bayes estimators.

E. Miscellaneous. Problems in several other areas of investigation related to the general problem of reliability assessment were to be studied. These include hazard rate estimation under censoring, prediction intervals, discrete failure models, simultaneous inference, and optimal design for treatment comparisons with a control.
3. **RESEARCH ACCOMPLISHMENTS**

In this section a substantive statement of the progress and significant accomplishments towards achieving the research objectives for this project outlined in section 2 will be given. The specific research papers referred to will be listed in section 4 of this report, which shows the cumulative results of the research project.

**A. Nonparametric estimation of quantiles under censoring.** Based on right-censored data from a lifetime distribution $F_0$, a kernel-type estimator of the quantile function $Q^0(p) = \inf\{t: F_0(t) \geq p\}$, $0 \leq p \leq 1$, was proposed by Padgett in paper number [5] listed in section 4. The estimator is defined by $Q_n(p) = n^{-1}\int_0^\infty Q^0(t)K((t-p)/h_n)dt$, which is smoother than the usual product-limit function $Q_n(p) = \inf\{t: F_n(t) \geq p\}$, where $F_n$ denotes the product-limit estimator of $F_0$. Under the random censorship model and general (but nonrestrictive) conditions on $h_n$, $K$, and $F_0$, it was shown in [5] that $Q_n(p)$ is strongly consistent. In addition, it was shown that an approximation, $Q_n^*$, to $Q_n$ is asymptotically equivalent to it with probability one. The general conditions for consistency included that $h_n \to 0$ as $n \to \infty$, that $K$ be a symmetric probability density with finite support, and that $F_0$ satisfy certain continuity conditions.

During this project, the estimator $Q_n(p)$ was studied further and in much more detail. In paper [7] it was shown that, under conditions similar to those required for strong consistency, $n^k(Q_n(p) - Q^0(p))$ is asymptotically normally distributed, provided $n^k h_n \to 0$. This condition on $h_n$ can be replaced by $n^k h_n \to 0$ by using a slightly different proof. By centering with the quantity $Q^0(p,h) = n^{-1}\int_0^\infty Q^0(t)K((t-p)/h_n)dt$ instead of $Q^0(p)$, an asymptotic normality result can be obtained without the above condition on the rate of convergence of $h_n$ to zero. Also, using this type of centering, the asymptotic normality of
the approximation \( Q^*_n \) was shown. These last two results are contained in paper number [9]. In addition, in [9], mean convergence of \( Q_n(p) \) and \( Q^*_n(p) \) were obtained and the mean square convergence was shown. In particular, an upper bound on the rate of mean square convergence of \( Q_n(p) \) to \( Q^0(p) \) was shown to be
\[
g(n,h_n) = o(h_n + h_n^{-5/4} + (\log n/n)^{3/4} + n^{-k}c(1-p)^2),
\]
where \( c \) is a constant.

An example of a bandwidth sequence satisfying \( g(n,h_n) \to 0 \) as \( n \to \infty \) is
\[ h_n = c_n n^{-\delta} \text{ for } 0 < \delta < 5/2 \text{ where } c_n > 0 \text{ is a bounded sequence.} \]
The value of \( h \) minimizing \( g(n,h) \) was found to be of the form \( h_n = Cn^{-5/6} \). This rate is not sharp, but seems to be close to optimal for large \( n \).

New asymptotic expressions for the mean squared errors of the product-limit quantile estimator and the kernel-type quantile estimator were presented in paper [14]. From these expressions a comparison of these two quantile estimators with respect to their mean squared errors was given. In particular, the comparison showed that for sufficiently large \( n \) and small enough \( h \), the kernel estimator's mean squared error was less.

Since the exact small-sample mean squared error has not been calculated, an expression for "optimal" \( h_n \) minimizing the mean squared error is not available. Hence, a study of the effect of varying the bandwidth was done by Monte Carlo simulations in paper [10]. Five common life distributions, two different censoring distributions, and three different kernel functions were used in the extensive simulations. The results indicate that at fixed \( p \) the kernel estimator \( Q_n(p) \) has smaller mean squared error than the product-limit quantile estimator \( \hat{Q}_n(p) \) for a range of values of the bandwidth \( h_n \). In addition, in [10] a method of selecting an "optimal" bandwidth value, in the sense of small estimated mean squared error, based on the bootstrap, was investigated. The results were consistent with the simulation study. The bootstrap was also used to obtain interval estimates of \( Q^0(p) \) after determining a value of \( h_n \) to use in computing \( Q_n(p) \).
In paper [19] a modification of the kernel quantile estimator was proposed in which the bandwidth \( h \) was a function of the censored data. The advantage of this estimator was that the data play a role in the degree of smoothing while retaining the desirable features of the kernel estimator. Consistency of the estimator was proven, and it was shown that a nearest-neighbor-type selection of the bandwidth gave a consistent estimator.

In [20], the smooth quantile estimation problem from censored data was approached by integrating a kernel density estimate to obtain a smooth estimate, \( \hat{F}_0 \), of the lifetime distribution. Then a solution \( x_p \) from \( \hat{F}_0(x_p) = p \) was obtained as an estimate of the \( p \)th quantile. Asymptotic properties of \( x_p \) were shown and some computational results were given.

Finally, in paper [17], for right-censored data a smooth alternative to the product-limit estimator as a nonparametric quantile estimator was proposed. This estimator, called a "generalized product-limit quantile estimator," was obtained by averaging appropriate subsample product-limit quantiles over all subsamples of a fixed size. Under the random censorship model and some conditions on the lifetime distribution, it was shown that the estimator was consistent and had the same asymptotic normal distribution as the product-limit quantile estimator. A small Monte Carlo simulation study showed that some values of the subsample size existed for which this "generalized" estimator performed better than the usual product-limit quantile estimator in the sense of estimated (from the simulations) mean squared errors.

B. Kernel density estimation from randomly right-censored data. The problem of asymptotically optimal selection of the bandwidth sequence for kernel density estimators under censoring was studied in paper number [13]. The usual kernel density estimator \( f_n(x) = h_n^{-1} \int_{-\infty}^{\infty} K((x-t)/h_n) d\hat{F}_n(t) \) and the Blum-Susarla estimator given by \( f_n^*(x) = \{nh_n \hat{H}_n(x)\}^{-1} \sum_{j=1}^{n} K((x-X_j)/h_n) I(\Delta_j=1) \)
were considered, where $H_n^*$ denotes an estimator of the censoring survival function, $(X_j, A_j)$, $j=1,...,n$, denote the right-censored data, and $I[.]$ denotes the indicator function of the event $[.]$. Two important contributions to the theory of bandwidth selection for kernel density estimators under right censorship were made in this paper. First, an asymptotic representation of the integrated squared error of each estimator in terms of easily understood variance and squared bias components is given. Second, it is shown that if the bandwidth is chosen by the data-based method of least squares cross-validation, then it is asymptotically optimal in a compelling sense. The optimal bandwidth is shown to be analogous to that for the case of no censoring. A by-product of the first part is an interesting comparison of the two kernel density estimators which shows that the difference between the two is typically not negligible.

C. Nonparametric estimation from step-stress accelerated life tests. In accelerated life testing, items are subjected to greater stress than that under the normal use conditions and, from the resulting failure data, an estimate of the lifetime distribution (or other inferences) under the normal use conditions is obtained. Most of the published work on inference from accelerated life tests require at least one of the following restrictions:

(i) that every item which is subjected to an overstress is observed until it fails or is removed (censored) under constant application of the stress, or

(ii) that the lifetime distribution of the items under every accelerated stress level is assumed to be known except for the underlying parameters. A more realistic approach allows the stress on an unfailed item to change at a preassigned test time. This approach is used in government defense studies and is referred to as step-stress testing.
In paper number [15] step-stress testing is considered from a nonparametric point of view when the observations have been arbitrarily right censored, relaxing both restrictions (i) and (ii) above. Thus, no assumptions are made about the particular form of the underlying lifetime distribution. Using a rescaling method, an estimator of the life distribution under the normal use conditions is obtained. The estimator is shown to be strongly consistent.

D. Linear empirical Bayes estimation of mean time to failure. In paper [11], it has been successfully established that not only the linear empirical Bayes estimator approaches the optimal linear estimator with probability one, but also in the quadratic mean with a rate of convergence inversely proportional to the number of problems considered. The same is true for the mean squared error of the estimators. Indeed, the limiting constant has been explicitly evaluated, which will provide a guideline for the consideration of a moderate number of problems. Effort will be continued to study the moderate sample properties of the linear empirical Bayes estimators, and to identify a natural and appropriate intermediate class of estimators.

Progress has also been made on an investigation into identification of a natural and appropriate intermediate class of estimators. Specifically, the class of quadratic estimators has been singled out and some preliminary properties have been shown. Further work needs to be done to obtain comprehensive and definitive results.

E. Miscellaneous. In other work, an invited paper [4] was written on nonparametric methods for hazard rate estimation from censored data. This paper first reviewed the available methods in the literature and then presented some new results on a kernel-type estimator of the hazard rate under random right censoring. The Koziol-Green (or proportional hazards) model was assumed for asymptotic results on the bias and mean squared error of the kernel estimator.
In paper number [8], a comparison was given of two prediction intervals for a future observation from the inverse Gaussian distribution. This distribution arises as a lifetime model which is an alternative to the lognormal distribution. One prediction interval is exact and the other is an approximate one. Surprisingly, the comparison showed that the approximate interval is superior in general, since it always yields a two-sided interval and has high coverage probability as well as small expected length.

In paper number [16] a regenerative sampling plan is developed for the sequential comparison of two populations having positive integral response. It is designed to be both an extension and an improvement of the play-the-winner rules for binary trials in the sense that a much wider variety of responses is allowed, the fraction of inferior selections approaches zero, and the play-the-winner rule is contained as a special case. Almost sure convergence and moment convergence in the pth order is studied for the fraction of inferior selections and for a maximum likelihood estimator of the mean response. A conditional test of hypothesis is given for the binary case. These results have applications in clinical trials and in quality control as well as in a variety of other sequential situations.

In paper number [18], a simple procedure is proposed to determine a sample size for estimating the mean weight of items in a problem of obtaining a large number of items. The situation is as follows: It is desired to obtain a large number, \( N_s \), of items for which individual counting is impractical, but one can demand a batch to weigh at least \( w \) units and hope that the number of items in the batch is close to the desired number \( N_s \). If the items have mean weight \( \theta \), it is reasonable to have \( w \) equal to \( \theta N_s \) when \( \theta \) is known. When \( \theta \) is unknown, one can take a sample of size \( n \), not bigger than \( N_s \), to estimate \( \theta \) by a good estimator \( \hat{\theta}_n \) and set \( w \) equal to \( \hat{\theta}_n N_s \). The proposed procedure determines the
sample size to be the integer closest to $pCN_s$, where $C$ is a function of the cost coefficients if the coefficient of variation $p$ is known. It is known to be optimal in some sense. If $p$ is unknown, a simple sequential procedure is proposed for which the average sample number is shown to be asymptotically equal to the optimal fixed sample size. When the weights are assumed to have a gamma distribution given $\theta$ and $\theta$ has a prior inverted gamma distribution, the optimal sample size in some sense can be found to be the nonnegative integer closest to $pCN_s + p^2A(pC-1)$, where $A$ is a known constant given in the prior distribution.

In paper [21], it was shown that the Gompertzian or Weibull (with increasing failure rate, IFR) distribution cannot arise from any mixture of exponential distributions. The proof of this used well-known results in reliability theory (but not well-known in the biology of aging). Therefore, if the population is Weibull (IFR) or Gompertzian, has bathtub-shaped failure rate, or has a failure rate that increases strictly over some interval, then there must be units that individually do not have exponential life distribution. This result is important in the analysis of reliability data.

4. CUMULATIVE LIST OF WRITTEN PUBLICATIONS

In this section, the research papers that have been written under this grant are listed. They are divided into three categories: In print, accepted for publication, and submitted. Copies of the manuscripts and reprints have been forwarded to the Program Manager as they were submitted or published.

A. In Print


B. Accepted for Publication


C. Submitted for Publication


[21] F. M. Guess and M. Witten, A population of exponentially distributed individual lifespans cannot lead to Gompertzian or to Weibull (with increasing mortality rate) dynamics. Submitted to *Science.*

5. **PROFESSIONAL PERSONNEL ASSOCIATED WITH THE RESEARCH EFFORT**

In addition to the principal investigator, W. J. Padgett, several co-investigators have been partially supported by this grant since June 1, 1984. John D. Spurrier was partially supported during the period from June 1, 1984, to May 31, 1985, and Kai F. Yu was partially supported from June 1, 1985, to May 31, 1987. During the summer of 1985, Diane T. McNichols, Department of Statistics, Virginia Polytechnic Institute and State U., was partially supported. Frank M. Guess was partially supported as a co-investigator from July 1, 1986, to May 31, 1987. Also, two graduate students have been supported, André M. Lubecke during the summer of 1984, and Y. L. Lio from August 16, 1984, to May 31, 1987. Ms. Lubecke completed the Ph.D. degree in August, 1985, and Mr. Lio will obtain the Ph.D. degree in August, 1987.

6. **INTERACTIONS**

The investigators attended several meetings and conferences and gave (invited and contributed) talks as follows:

i) W. J. Padgett, J. D. Spurrier attended the Conference on Reliability and Quality Control, University of Missouri, Columbia, MO, June 4-8, 1984.

iii) J. D. Spurrier, "The training of statistical consultants at the University of South Carolina" (invited). Joint Statistical Meetings, Philadelphia, PA, August 1984.


vi) W. J. Padgett was the invited discussant for the Session on Reliability and Life Testing, Spring Regional Meeting of the Biometric Society (ENAR) and ASA in Raleigh, NC, March 25-27, 1985.


viii) W. J. Padgett gave an invited talk on "Nonparametric density and failure rate estimators when samples are censored" at the SREB-ASA Summer Research Conference on Statistics, Boone, NC, June 16-21, 1985.


xiv) K. F. Yu attended the NSF-CBMS Conference on Nonparametric Priors, University Park, PA, June 3-7, 1985.


xvi) W. J. Padgett gave an invited talk on "Some results for kernel-type quantile estimators from right-censored data" at the U.S. Army


xiv) W. J. Padgett gave an invited talk on "Asymptotic results for kernel quantile estimators under censoring" at the Charlotte Mathematics Conference (AMS/MAA), Charlotte, NC, on October 17-18, 1986.


xxiii) F. M. Guess gave an invited talk on "Mean residual life: theory and applications" at the Southeastern Regional American Mathematical Society meeting, Charlotte, NC, October 18, 1986.

xxiv) K. F. Yu and W. J. Padgett attended Joint Statistical Meetings, Chicago, August 17-21, 1986, and Dr. Yu presented a paper entitled "A note on a renewal theorem for a moving average process."

xxv) F. M. Guess and K. F. Yu attended the 9th Annual Meeting of the S.C. Chapter of ASA. April 10, 1987, in Columbia, SC, and were judges of the student presentations.

7. INVENTIONS, PATENT DISCLOSURES, AND APPLICATIONS STEMMING FROM THE RESEARCH PROJECT

No inventions or patents have stemmed from this research.

The results reported in Section 4 have wide application in the estimation and assessment of reliability and maintenance of military equipment. The various nonparametric procedures with censored data developed in this project allow accurate estimation of median and mean lifetime, percentiles, densities, survival probability, and other lifetime characteristics without assuming
particular forms of the life distribution. In particular, the kernel-type quantile estimator will provide more accurate estimates of the median lifetime and smaller percentiles of a piece of equipment than the usual product-limit estimators.

8. OTHER PROFESSIONAL ACTIVITIES

During this reporting period, June 1, 1984, to May 31, 1987, the investigators have been involved in numerous other professional activities that are intimately related to the research efforts on this grant. Professor Padgett has refereed 27 manuscripts for Technometrics, the Journal of the American Statistical Association, Communications in Statistics, Statistics and Probability Letters, Pakistan Journal of Statistics, Biometrika, IEEE Transactions on Reliability, Journal of Multivariate Analysis, Journal of Statistical Planning and Inference, Operations Research, and the Journal of the University of Kuwait (Science). Professor Yu refereed six articles for the Annals of Statistics, Sequential Analysis, Statistics and Probability Letters, JASA, and the Journal of Statistical Planning and Inference. Professor Guess has refereed one article for each of the Journal of the American Statistical Association and the Naval Research Logistics Quarterly. In addition, five papers were reviewed for the Zentralblatt für Mathematik and 21 were reviewed for the Mathematical Reviews. Also, four research proposals were reviewed for the APOS, four for the ARO, and one for the NSF.

Professor Padgett was a member of the International Editorial Board of the Communications in Statistics-Theory and Methods and was an Associate Editor of Naval Research Logistics Quarterly and the Journal of Statistical Computation and Simulation. In addition, he was the Program Co-Chair for the SRCOS-ASA Summer Research Conference in Statistics held in Mobile, AL, on June 15-20,
1986. He also chaired an invited paper at the Spring Statistics Meetings in Atlanta, GA, in March, 1986. Professor Spurrier was a member of the International Editorial Board of the Communications in Statistics - Statistical Reviews and was an Associate Editor of Technometrics.
END
10-81
DTIC