The Jackknife, the Bootstrap, and Censored Data
A Review and Simulation Study

Michael S. Steinbach
University of Minnesota

Technical Report No. 411

November 1982

NSF Grant MSC 7904558.

Principal Investigator, D.V. Hinkley

1. Introduction

In survival studies the experimenter is sometimes unable to observe the lifetime X° of every subject in the experiment because the study may end while the subject is still alive, or the subject may leave during the study. In both cases, the experimenter will then know a lower bound Y on the subject's lifetime. It is conventional to refer to Y as a censored observation, or a censoring time. Y may or may not be random relative to X°. In either case, what is known as the censored data problem consists of attempting to make inferences about the distribution of the lifetime X° given that what is observable is $X = \min(X^\circ, Y)$.

It is rarely possible to characterize the distribution of X° as being a member of a known parametric family of distributions. Consequently, it is of considerable interest to obtain a nonparametric estimate of F°, the c.d.f. of X°, or equivalently S°, the survival function of X°, where $S^\circ(t) = 1 - f^\circ(t) = \Pr(X^\circ > t)$. There are a number of established procedures for estimating S°, but this paper will consider only the Kaplan-Meier method (defined explicitly in Section 4).

If a characteristic θ of the lifetime distribution, such as the mean or median, is of interest, then θ can be characterized as a functional of F° or S°, say $\theta = t(F^\circ)$. Then given $\hat{F}^\circ$, an estimate of F°, the natural estimate of θ is simply $\hat{\theta} = t(\hat{F}^\circ)$. In general, however, it is not a simple task to obtain estimates for the bias and variance of $\hat{\theta}$, or to obtain confidence intervals for θ. The jackknife and bootstrap are two related nonparametric methods which are applicable to such problems.

In this paper we attempt to evaluate, in a limited way, the performances of the jackknife and bootstrap with respect to inference about the mean lifetime in a censored data problem. For these purposes, underlying distributions were assumed for the survival and censoring times and a number of Monte Carlo simulations were done. Since the distributions are known to us, it is then possible to study parametric confidence intervals for the mean, which represent a "control" against which to judge the non-parametric methods.

2. The Jackknife

The jackknife is a very general numerical procedure which can be used to give estimates of the bias and variance of a statistic $\hat{\theta}$. If a normal approximation is reasonable for $\hat{\theta}$, it is then possible to construct approximate confidence intervals for θ, the parameter estimated by $\hat{\theta}$. In this section we outline the general jackknife method, together with some of the supporting theory.

We assume here that θ is a functional map $t(F)$ of the underlying distribution F, i.e. $t:C \to R$ where C is a set of one-dimensional c.d.f.'s, $F \in C$. As an example, consider the case where θ, the distributional characteristic of interest, is the mean. For any c.d.f. F, the mean is uniquely defined by the equation, $\theta = \int_{-\infty}^{\infty} z dF(z)$, if the integral exists. Let C be the set of all c.d.f.'s for which the mean is defined. Clearly, the previous equation defines a functional t from C to R such that $\theta = t(F)$, $\forall\ F \in C$.

If $\hat{F}$ is an estimate of F, then $\hat{\theta} = t(\hat{F})$ is an estimate of θ. Such a definition of $\hat{\theta}$ implies that $\hat{\theta}$ is consistent if $\hat{F}$ is consistent, provided t is a continuous function in an appropriate sense. Furthermore, in the nonparametric context, this definition of $\hat{\theta}$ has robustness of validity since it does not depend on a model assumption.

We describe next some assumptions and notation pertinent to the jackknife. Let $x_1,\ldots,x_n$ be a random sample of size n drawn from F, where $F \in C$ and $\theta = t(F)$. Let $\hat{F}$ be the sample c.d.f. of $x_1,\ldots,x_n$, where $\hat{F} \in C$ and $\hat{\theta} = t(\hat{F})$. Define $\hat{F}_{(j)}$ to be the corresponding estimate of F derived from the subsample $x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n$, i.e. the subsample with x_j omitted. Then, assuming that $\hat{F}_{(j)} \in C$, $1 \leq j \leq n$, let $\hat{\theta}_j = t(\hat{F}_{(j)})$ be the jth subsample estimate of θ.

The first step in the jackknife procedure is the calculation of $\hat{\theta}_j$, $1 \leq j \leq n$. The next step is the calculation of the sample influence values $\tilde{I}_j$, $1 \leq j \leq n$, defined by

$$\tilde{I}_j = (n-1)\{t(\hat{F}) - t(\hat{F}_{(j)})\} \tag{2.3}$$
$$= (n-1)(\hat{\theta} - \hat{\theta}_j).$$

It turns out that $\tilde{I}_j$ is ultimately related to the theoretical influence function which is defined below.

Let b represent the bias of $\hat{\theta} = t(\hat{F})$, where $b = E(\hat{\theta}) - \theta = E[t(\hat{F})] - t(F)$. Then, $\tilde{b}$, the jackknife estimate of b, is

$$\tilde{b} = -\left(\sum_{j=1}^{n} \tilde{I}_j\right)/n = -\bar{\tilde{I}}. \tag{2.2}$$

Subtraction of $\tilde{b}$ from $\hat{\theta}$ gives the jackknife estimate of θ,

$$\tilde{\theta} = \hat{\theta} - \tilde{b} = \hat{\theta} + \bar{\tilde{I}}. \tag{2.3}$$

Let V represent the variance of $\hat{\theta} = t(\hat{F})$. Then $\tilde{V}$, the jackknife estimate of V, is given in terms of the $\tilde{I}_j$ by

$$\tilde{V} = \sum_{j=1}^{n} (\tilde{I}_j - \bar{\tilde{I}})^2/\{n(n-1)\}. \tag{2.4}$$

A full description of the theory supporting $\tilde{b}$ and $\tilde{V}$ is beyond the scope of this paper, but we can outline the basic element. The main ideas flow from an expansion of $t(\hat{F})$ using a generalization of Taylors' method, known as the von Mises expansion.

If G and F are c.d.f.'s and if t is a functional, then the von Mises expansion for t(G) in terms of t(F) is given by (6) below, where D_1 and D_2 are respectively, the first and second order differential operators, and where the x and y in t(x;F) and t(x;y;F) are included in the functional notation to indicate the appropriate dummy variable with respect to which D_1 and D_2 operate.

$$t(G) = t(F) + \int D_1 t(x;F)\,d(G-F)(x) + \tfrac{1}{2}\iint D_2 t(x;y;F)\,d(G-F)(x)\,d(G-F)(y) + \ldots \tag{2.5}$$

If $\hat{F}$ is the sample c.d.f. of $x_1,\ldots,x_n$, then substituting $G = \hat{F}$ in (2.5) and rearranging terms yields (2.6) below, with I_t and Q_t defined in (2.7) and (2.8), respectively.

$$t(\hat{F}) = t(G) + \int I_t(x;F)\,d\hat{F}(x) + \tfrac{1}{2}\iint Q_t(z;j;F)\,d\hat{F}(x)\,d\hat{F}(y) + \ldots \tag{2.6}$$

$$I_t(x;F) = D_1 t(x;F) - \int D_1 t(u;F)\,dF(u) \tag{2.7}$$

$$Q_t(x;y;F) = D_2 t(x;y;F) - \int D_2 t(u;y;F)\,dF(u) - \int D_2 t(x;v,F)\,dF(v) + \iint D_2(u;v;F)\,dF(u)\,dF(v) \tag{2.8}$$

I_t is the influence function and will be discussed below. For future reference, note that $E\, I_t(x;F) = E_{x\neq y}[Q_t(x;y;F)] = 0$. Since $\hat{F}$ assigns mass $1/n$ to x_j, $1 \le j \le n$, (2.6) becomes

$$t(\hat{F}) = t(F) + \frac{1}{n}\sum_{j=1}^{n} I(x_j;F) + \frac{1}{2n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} Q_t(x_j,x_k;F) + \ldots \tag{2.9}$$

The theoretical basis of the jackknife is the approximation of $t(\hat{F})$ by the first three terms of (2.6) or (2.9), where we assume that Q_t exists and is continuous in F.

To understand the influence function $I_t(x;F)$ and its relationship to $\tilde{I}_j$, it is helpful to rewrite (2.7) as

$$I_t(x;F) = \frac{d}{dt}\, t\big|(1-\epsilon)F + \epsilon\delta x\big|_{\epsilon=0}, \tag{2.10}$$

where δ_x is the c.d.f. defined by $\delta_x(y) = 0$, $\forall y < x$ and $\delta_x(y) = 1$, $\forall y \geq x$ (i.e. δ_x puts unit mass at x). From (2.10) it can be seen that $I_t(x;F)$ is a measure of the change in t(F) if the mass at x is changed by an infinitesimal amount. Recall that $\tilde{I}_j$ is a measure of the change in $t(\hat{F})$ when the mass at x_j is changed, i.e. when x_j is omitted. There is a strong similarity in the definitions of $I_t(x;F)$ and $\tilde{I}_j$. Indeed, under suitable smoothness assumptions,

$$(2.11) \quad \tilde{I}_j = I_t(x_j;F) + o(n^{-1}).$$

We can now give a rough, but simple justification of (2.2) and (2.3), the formulas for $\tilde{b}$ and $\tilde{V}$, respectively. Rewrite (2.9) as

$$(2.12) \quad t(\hat{F}) - t(F) = \sum_{j+1} I(x_j,F) + \frac{1}{2n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} Q(x_j;x_k;F) + \ldots .$$

Taking the expectation on both sides yields

$$\text{bias } [t(F)] = E[t(F)] - t(F) = \frac{1}{2n} E[Q(x;x;F)] + \ldots .$$

It can be shown that $E[\tilde{I}_j] = -\frac{1}{2}E[Q(x;x;F)] + \ldots$, $1 \leq j \leq n$. Therefore, if $\tilde{I}_\cdot$ is used to estimate $E[\tilde{I}_j]$, (2.2) follows immediately. If we take the variance on both sides of (2.12) we get $\text{var}\,(t(\hat{F})) = \frac{1}{n}\text{var}[I(x;F)] + \ldots$. But (2.11) implies $\text{var}\,[I(x;F)] \approx \text{var}\,[\tilde{I}_j]$. If $\frac{1}{n-1} \sum_{j=1}^{n} (\tilde{I}_j - \bar{\tilde{I}}_\cdot)^2$ is used to estimate $\text{var}\,[\tilde{I}_j]$, (2.3) follows immediately.

Based on the first two terms of (2.9), application of the central limit theorem shows that $(t(\hat{F}) - t(F))/\sqrt{\text{Var}\,[t(\hat{F})]}$ is asymptotically N(0,1), where $\text{var}\,[t(\hat{F})] = \frac{1}{n}\,\text{var}\,[I(x;F)]$, asymptotically. If it can further be assumed that $\tilde{V}$ is a consistent estimator of $\text{var}[t(\hat{F})]$, then $(t(\hat{F}) - t(F))/\sqrt{\tilde{V}}$ is asymptotically N(0,1). Thus, an approximate 100(1-p)th percent confidence interval for $\theta = t(F)$ is given by

$$(2.13) \quad (t(\hat{F}) - z_{p/2}\sqrt{\tilde{V}}\,,\ t(\hat{F}) + z_{p/2}\sqrt{\tilde{V}}\,),$$

where $z_{p/2}$ is the 100 p/2th percentile of the N(0,1) distribution.

A comment on transformations is appropriate at this point. Sometimes, especially when n is small, the normal approximation for $t(\hat{F}_n)$ referred to above is not very accurate. Consequently, the normal confidence interval (2.13) for θ may not be very accurate in the sense that (a) the actual coverage rate may be considerably different from the desired coverage rate, and/or (b) the probability of a right error (the confidence interval falls to the right of θ) may be substantially different than the probability of a left error. However, it may be possible to find a monotone continuous function (transformation) f, such that $f(t(\hat{F}))$ is more amenable to accurate normal approximation; the same asymptotic theory will apply. Hence, an approximate normal confidence interval for $f(\theta)=f(t(F))$ can be constructed which will be relatively accurate in terms of (a) and (b). To be specific, a 100(1-p)th percent confidence interval for $f(\theta)$ is given by

$$(2.14) \quad (f(t(\hat{F})) - z_p\sqrt{\tilde{V}_f},\ f(t(\hat{F})) + z_p\sqrt{\tilde{V}_f}),$$

where $\tilde{V}_f$ is calculated as in (2.4) given that $\tilde{I}_j^f = (n-1)\{f(t(\hat{F})) - f(t(F_{(j)}))\}$. The confidence interval for $f(\theta)$ can then be inverted to form a confidence interval for θ which is accurate in the sense of (a) and (b). Finally, note that except for a few well analyzed cases, the best choice of f is usually unclear.

3. The Bootstrap

The bootstrap is another procedure especially well suited to dealing with nonparametric problems. Like the jackknife, it can be used to give estimates of the bias and variance of a statistic $\hat{\theta} = t(\hat{F})$. Furthermore, the bootstrap can give confidence intervals for θ which do not require any assumption of normality.

The basic idea of the bootstrap is very simple. Suppose, for example, that we wish to calculate a quantitative property of the distribution of $\hat{\theta} = t(F)$, such as $E(\hat{\theta}) - \theta$ or $Pr(\hat{\theta} \leq a)$. If this property is denoted by $N_t(F)$, then we estimate it by $N_t(\hat{F})$, the corresponding property of $\hat{\theta}^* = t(\hat{F})$. More explicitly, to estimate $E(\hat{\theta}) - \theta$, for example, we use

$$N_t(\hat{F}) = E(\hat{\theta}^*) - \hat{\theta}.$$

The expectation on the right hand side may be calculable theoretically, but this is usually not possible. We then resort to Monte Carlo simulation of the expectation. It is the estimation of $N_r(F)$ by a simulation estimate of $N_r(\hat{F})$ which is the actual bootstrap procedure.

The details of the bootstrap are as follows: N pseudo-random "bootstrap" samples of size n are generated from $\hat{F}$. If $\hat{F}$ is the sample c.d.f. of the original sample $X_1, \ldots X_n$, then generating a bootstrap sample $X_1^*, \ldots, X_n^*$ amounts to sampling n times independently and with replacement from $X_1, \ldots, X_n$. (The asterisk superscript is used to distinguish bootstrap quantities from the corresponding original sample quantities, eg. $\hat{\theta}$ and $\hat{\theta}^*$.) The value of $\hat{\theta}^*$ is computed for each of the N bootstrap samples, yielding a single sample size N of $\hat{\theta}^*$ values which can be used to estimate $N_t(\hat{F})$. For example, if $N_t(\hat{F}) = E(\hat{\theta}^*)$, the sample average is the appropriate estimate.

Let $\hat{\theta}^*_j$ be the value of $\hat{\theta}^*$ computed from the jth bootstrap sample. Then the bootstrap estimates of the bias and variance of $\hat{\theta}$ are, respectively,

$$(3.1) \quad \hat{b}^* = \sum_{j=1}^{N} \hat{\theta}^*_j / N - \hat{\theta};$$

$$(3.2) \quad \hat{v}^* = \sum_{j=1}^{N} (\hat{\theta}^*_j - (\sum_{i=1}^{n} \hat{\theta}^*_i / N))^2 / N - 1.$$

If a normal approximation for $\hat{\theta}$ is reasonable, an approximate $100(1-p)$th percent confidence interval for θ is

$$(3.3) \quad (\hat{\theta} - z_{p/2} \sqrt{\hat{v}^*} \, , \, \hat{\theta} + z_{p/2} \sqrt{\hat{v}^*} \,).$$

As mentioned, we can construct confidence intervals for θ which do not require an assumption of normality for $\hat{\theta}$. To this end, assume that $\hat{\theta} - \theta$ is pivotal, i.e. $\Pr(\hat{\theta} - \theta \leq q_p | \theta) = p \ \forall \theta$. Consequently, $\Pr(\theta \geq \hat{\theta} - q_p) = p \quad \forall \theta$, and $\hat{\theta} - z_p$ is a 100 pth percent lower confidence bound for θ. $\hat{\theta}$ is known and z_p can be estimated via the bootstrap as the 100 pth percentile of $(\hat{\theta}^* - \hat{\theta}) = y_p - \hat{\theta}$, where y_p = 100 pth percentile of $\hat{\theta}^*$. Therefore, $\hat{\theta} - q_p \approx 2\hat{\theta} - y_p$, and an approximate 100(1-p)th percent confidence interval for θ is given by

$$(3.4) \quad (2\hat{\theta} - y_{p/2}, \ 2\hat{\theta} - y_{(1-p/2)}).$$

4. C.D.F. and Mean Lifetime Estimation for Censored Data

For the purposes of this paper, we need to consider the censored data problem in more detail. This requires a brief sketch of the notation and basic situation. The random variable of interest is X°, the lifetime of an experimental subject. We shall assume that the censoring time Y is also a random variable. In an actual experiment, the observed quantities are $/X = \min(X^\circ, Y)$ and $d = \begin{cases} 0 & \text{if } X=Y \\ 1 & \text{if } X=X^\circ \end{cases}./$ F, F°, and G are the c.d.f.'s of X, X°, and Y, respectively. Let the subscript j refer to the jth subject in a random sample of size n. $X_1^\circ,\ldots,X_n^\circ$ are i.i.d. with c.d.f. F , and $Y_1,\ldots,Y_n$ are i.i.d. with c.d.f. G. Furthermore, Y_j is assumed to be independent of X_j. The n observations in a sample are the pairs $(X_1,d_1),\ldots,(X_n,d_n)$, where X_j and d_j have the obvious meanings. As a result of the previous assumptions, the X_j are i.i.d. with c.d.f. F. OX_j will denote the jth order statistic of $X_1,\ldots,X_n$. If $OX_j = X_k$, then $d_{[j]}$ is the corresponding indicator of censoring, i.e. $d_{[j]} = d_k$.

It is important in censored data problems to obtain a nonparametric estimate of F°. This paper uses the Kaplan-Meier (K-M) procedure to get such an estimate. For convenience, the following discussion uses the survival function $S^\circ = 1 - F^\circ$ instead of F°. To begin, suppose that $ox_1,\ldots,ox_n$ are the observed values of $OX_1,\ldots,OX_n$. For convenience, also assume that the ox_j are distinct. Define $p_1 = PR(X^\circ > ox_1)$, and for $j > 1$ define $p_j = Pr(X^\circ > ox_j \,|\, X^\circ > ox_{j-1})$. Then, $Pr(X^\circ > OX_k) = \prod_{j=1}^{n} p_j$. The K-M method estimates p_j by $/\hat{p}_j = (\#$ survivors at time $ox_j)/(\#$ survivors just before $ox_j)/ = 1 - \#$ deaths at $ox_j/\#$ survivors just before ox_j. Since the

ox_j are distinct, $\hat{p}_j = 1 - \frac{1}{n-j+1}$ if $d_{[j]} = 1$, and $\hat{p}_j = 1$ if $d_{[j]} = 0$. Note that no information is available on the behavior of S between ox_{j-1} and ox_j. Consequently, $\hat{S}(t) = \hat{Pr}(X_0 > t) = \hat{Pr}(X^\circ > ox_{j-1})$ if $ox_{j-1} < t < ox_j$. The preceeding facts are summarized in equation (4.1) which gives the formula for $\hat{S}^\circ$, the K-M estimate of S.

$$(4.1) \qquad \hat{S}^\circ(t) = \prod_{j:\, ox_j < t} \left(1 - \frac{1}{n-j+1}\right)^{d_{[j]}} .$$

If $d_{[n]} = 0$, $\hat{S}(t)$ never reaches 0. To prevent this unrealistic and awkward property of $\hat{S}$, we define $\hat{S}(t) \equiv 0$, for $t \geq ox_n$. Finally, note that $\hat{F}^\circ = 1 - \hat{S}^\circ$.

The main focus of this paper will be on the mean μ of F°. We express the mean of F° in the form $t(F^\circ) = \int_0^\infty x\, dF^\circ(x)$, assuming the integral exists, and thence define the K-M estimate of μ as $\hat{\mu} = t(\hat{F}^\circ) = \int_0^\infty x\, d\hat{F}^\circ(x)$. In terms of the observed OX_j and $\hat{S}^\circ$,

$$(4.2) \qquad \hat{\mu} = \sum_{j=1}^{n} (ox_j - ox_{j-1})\, \hat{S}^\circ(ox_{j-1}^-)$$

where $ox_0 \equiv 0$, and $\hat{S}(ox_0) \equiv 0$.

The K-M estimators $\hat{\mu}$ and $\hat{S}^\circ$ have a number of nice properties. $\hat{S}^\circ$ is (a) the maximum likelihood estimator (m.l.e.) of S, and (b) a consistent estimator of S. Properties (a) and (b) of $\hat{S}^\circ$ and the functional nature of $\hat{\mu}$, $\hat{\mu} = t(1 - \hat{S}^\circ)$, imply that (1) $\hat{\mu}$ is the m.l.e. of μ, and (2) $\hat{\mu}$ is consistent. It can also be shown that $\hat{\mu}$ is asymptotically normal for reasonable F° and G. Finally, not that $\hat{S}^\circ$ is a m.l.e. on an infinite-dimensional parameter space, and thus, the customary properties of a m.l.e. do not necessarily hold for either $\hat{S}^\circ$ or $\hat{\mu}$.

5. Parametric Estimation of the Mean Lifetime

Since F° and G are generally unknown, parametric methods of statistical analysis for the censored data problem are considered here basically as a reference point for the jackknife and bootstrap. In particular, we will consider two confidence interval procedures for μ which are based on the likelihood of μ given $(X_1,d_1),\ldots,(X_n,d_n)$. To this end, we assume that X° and Y are exponentially distributed with $E(X^\circ) = \mu = 1/\lambda_1$, and $E(Y) = \theta = 1/\lambda_2$, i.e. $F^\circ(t) = 1 - e^{-\lambda_1 t}$, and $G(t) = 1 - e^{-\lambda_2 t}$.

Let $L(\lambda_1)$ be the likelihood of $\lambda_1 = 1/\mu$. Given $(X_1,d_1),\ldots,(X_n,d_n)$. By definition, $L(\lambda_1) = \prod_{j=1}^{n} \lim_{\Delta\to 0} \Pr(x_j \le X_j < x_j + \Delta)/\Delta$. To evaluate $L(\lambda_1)$, note that

$$\lim_{\Delta\to 0} \Pr(x_j \le X_j < x_j + \Delta,\ d_j = 1)/\Delta$$

$$= \lim_{\Delta\to 0} \Pr(x_j \le X_j^\circ < x_j + \Delta,\ Y_j > X^\circ)/\Delta$$

$$= \lim_{\Delta\to 0} \Pr(x_j \le X_j^\circ < x_j + \Delta)\ \Pr(Y_j > x_j + \Delta)$$

$$= \lim_{\Delta\to 0} \lambda_1 e^{-\lambda_1 x_j}\ \Delta\ e^{-\lambda_2 (x_j + \Delta)/\Delta}$$

$$= \lambda_1 e^{-\lambda_1 x_j}\ e^{-\lambda_2 x_j}\ .$$

Similarly, $\lim_{\Delta\to 0} \Pr(x_j \le X_j \le x_j + \Delta,\ d_j = 0) = \lambda_2 e^{-\lambda_2 x_j}$. Therefore, $L(\lambda_1)$ $= \lambda_1^r e^{-\lambda_1 \sum_{j=1}^{n} x_j}$, where r = # of uncensored data points. The log-likelihood is

$$(5.1) \quad L(\lambda_1) = \log(L(\lambda_1)) = r \log(\lambda_1) - \lambda_1 \sum_{j=1}^{n} x_j .$$

The two methods that will be used for calculating confidence intervals are (i) normal intervals contered on the m.l.e., and (ii) likelihood ratio intervals. Recall that the m.l.e. of λ_1 is the solution to $\partial L/\partial\lambda_1 = 0$, i.e. $\hat{\lambda}_1 = r/\sum_{j=1}^{n} x_j$. If now, $I_{\lambda_1} = -\partial^2 L/\partial\lambda_1^2\,|_{\hat{\lambda}_1}$, the observed Fisher information with respect to λ_1, then for large n, $I_{\lambda_1}(\hat{\lambda}_1 - \lambda_1)^2 \approx \chi_1^2$. It is likely that this result is more accurate for the transformed parameter $q = \log(\mu) = -\log(\lambda_1)$,

for which

$$I_q = -\frac{\partial^2}{\partial q^2}(L(\lambda_1))\Big|_{\hat{q} = \log(\hat{\lambda}_1)}$$

$$= \left(\frac{\partial \lambda_1}{\partial q}\right)^2 \Big|_{\hat{q}} \left(-\frac{\partial^2}{\partial \lambda_1^2} L(\lambda_1)\right)\Big|_{\lambda_1} = r.$$

Hence, an approximate 100(1-p) percent confidence interval for q will be

$$(5.2) \quad (\hat{q} - z_{p/2}/\sqrt{r}, \hat{q} + z_{p/2}/\sqrt{r}).$$

The second (invariant) method is based on the likelihood ratio, using the fact that $2\{L(\hat{\lambda}_1) - L(\lambda_1)\} \approx \chi_1^2$ for large n. Thus,

$$(5.3) \quad \{q: 2[L(\hat{q}) - L(q)] \leq x_1^2(p)\}$$

is an approximate 100(1-p)% confidence interval for q. To express (5.3) in terms of sample quantities, note that $L(q) = -rq - e^{-q}\sum_{j=1}^{n} x_j$, $e^{\hat{q}} = e^{\log(\Sigma x_j/r)}$ $= \sum_{j=1}^{n} x_j/r$, and hence, $L(q) = -rq - re^{\hat{q}-q}$. Thus, $2[L(\hat{q}) - L(q)] = 2r(y + e^{-y} + 1)$, where $y = q - \hat{q}$. Hence an approximate 100(1-p) percent confidence interval for q is

$$(5.4) \quad (y_1 + \hat{q}, y_2 + \hat{q}),$$

where $y_1 < y_2$ are the solutions to $2r(y + e^{-y} + 1) = \chi_1^2(p)$.

6. Applying the Jackknife to Censored Data

It is straightforward to use the jackknife procedure to estimate the bias and variance of $\hat{\mu}$. With the assumption that $\hat{\mu}$ is approximately normal, it is also possible to give approximate confidence intervals for μ. However, it is reasonable to believe that $\log(\hat{\mu})$ is more nearly normal than $\hat{\mu}$ in finite samples; e.g. $\log(\hat{\mu})$ can be negative while $\hat{\mu}$ cannot. Therefore, it would also be desirable to do a jackknife analysis of $\hat{\mu}$ on the log scale. In what follows, the quantities related to a log scale analysis will be denoted by an L subscript or superscript.

All the formulas needed for the jackknife analysis of $\hat{\mu}$ and $\log(\hat{\mu})$ have already been given. Equation (4.2) is the formula for calculating $\hat{\mu} = t(\hat{F}^\circ)$, given the estimated survival function, $\hat{S}^\circ = 1 - \hat{F}^\circ$. Equation (4.1) is the formula for calculating $\hat{S}$, given the observed data $(X_1,d_1),\ldots,(X_n,d_n)$. Equations (4.1) and (4.2) can also be used to calculate $\hat{\mu}_j = t(\hat{F}^\circ_{(j)})$ by simply omitting (X_j,d_j) and using n-1 instead of n. However, in practice it is computationally more efficient to calculate the $\hat{\mu}_{(j)}$ by the following algorithm than to calculate each estimate by an application of (4.1) and (4.2).

Algorithm: Let M be the $n \times n$ matrix with elements $M_{i1} = 1$, $M_{i,i+1} = 1$; $M_{ij} = \left(\frac{1}{1-n-j-1}\right)^{d(j-1)}$, $2 \le j \le i$; $M_{ij} = \left(1 - \frac{1}{n-j-2}\right)^{d(j-2)}$, $j > i+1$; $M_{nn} = 0$ with $d(0) = 0$. Then $\hat{\mu}(j) = \sum_{j=1}^{n} (X_{(j)} - X_{(j-1)}) \prod_{k=1}^{j} M_{ik}$, where $\hat{\mu}_{(j)}$ is the $\hat{\mu}_j$ which results from omitting $X_{(j)}$. The formulas for $\tilde{b}$ and $\tilde{V}$, the jackknife estimates of the bias and variance of $\hat{\mu}$ are given by (2.2) and (2.3), where $\tilde{I}_j$ is as defined by (2.1). An approximate 100(1-p) percent confidence interval is as given by (2.13).

If $\tilde{I}_j^L = (n-1)\{\log(\hat{\mu}) - \log(\hat{\mu}_{(j)})\}$ is taken as a starting point, then the jackknife analysis of $\log(\hat{\mu})$ is completely analogous to the jackknife analysis of $\hat{\mu}$. The symbols $\tilde{b}_L$ and $\tilde{V}_L$ will represent the jackknife estimate of the bias and variance of $\log(\hat{\mu})$.

7. Applying the Bootstrap to Censored Data

The application of the bootstrap procedure to the censored data problem requires a bit of care. The previous description of the bootstrap involved only a single underlying distribution, whereas the censored data problem has two underlying distributions, F° and G. The major effect of this difference is in the manner in which bootstrap samples are generated.

The most general way in which to generate a bootstrap sample for the censored data problem is as follows. Obtain F° and G, the K-M estimates of F and G, from the given sample $(X_1,d_1),\ldots,(X_n,d_n)$. Generate an artificial sample of independent pairs, where $X_j^{\circ *} \sim \hat{F}^\circ$, and $Y_j^* \sim \hat{G}$, and where X_j^* and Y_j^* are independent. Then $(X_1^*,d_1^*),\ldots,(X_n^*,d_n^*)$ is the desired bootstrap sample, where $X_j^* = \min(X_j^{\circ *},Y_j^{\circ *})$ and $d_j^* = \begin{cases} 1, & X_j^* = X_j^{\circ *} \\ 0, & X_j^* = Y_j^* \end{cases}$. If Y is a random variable, as in this paper, an equivalent and simpler way to generate a bootstrap sample is the following: Obtain $(X_1^*,d_1^*),\ldots,(X_n^*,d_n^*)$ by sampling n times randomly and with replacement from $(X_1,d_1),\ldots,(X_n,d_n)$.

To understand this equivalence we consider the m.l.e.'s of $S = 1 - F$, $S^\circ = 1 - F^\circ$, and $R = 1 - G$. $\hat{S}$, the m.l.e. of S, is as defined in (4.1). The K-M estimate $\hat{R}$, the m.l.e. of $R = 1 - G$, is given by $\hat{R}(t) = \prod_{j:x_{(j)}<t} (1 - \frac{1}{n-j+1})^{1-d_{(j)}}$. If a, b, c are independent random variables having c.d.f.'s A, B, and C, and if $a = \min(b,c)$, then it is easily shown that $A = 1 - (1-B)(1-C)$, i.e. $(1-A) = (1-B)(1-C)$. From this we conclude that (1) $\hat{S} = \hat{S}^\circ\hat{R}$, i.e. $\hat{S}$ is the nonparametric m.l.e. of $S = 1 - F$, and (2) $\hat{S}$ is the c.d.f. of X_j^*, $1 \le j \le n$. The definitions of $\hat{S}$ and $\hat{R}$ imply

$$S(t) = \prod_{j:X_{(j)}<t} (1 - \frac{1}{n-j+1}) = \frac{n-k_t}{n},$$

where k_t is the largest j such that $X_{(j)} \leq t$. In other words, $\hat{S}(t)$ assigns probability $\frac{1}{n}$ to each X_j, $1 \leq j \leq n$.

$\hat{F}^\circ$ and $\hat{G}$, like $\hat{F}$, assign positive probability to only a finite number of points - some subset of $X_1,\ldots,X_n$. $\hat{F}^\circ$ assigns positive probability only to those X_j for which $d_j = 1$, while $\hat{G}$ assigns positive probability only to those X_j for which $d_j = 0$. Thus, if $X_j^* = X_j^{\circ *} = X_i$, $d_j = 1$ by definition and $d_i = 1$, and likewise, if $X_j^* = Y_j^* = X_i$, $d_j^* = 0$ by definition, and $d_i = 0$. Therefore, the only possible values of (X_j^*, d_j^*) in the first method are $(X_1,d_1),\ldots,(X_n.d_n)$. But $\Pr\{(X_j^*,d_j^*) = (x_i,d_i)\} = \Pr(X_j^* = X_i) = \frac{1}{n}$ by the previous paragraph, and we have shown that the first and second methods are equivalent. The rest of the bootstrap procedure for the consored data problem is as originally described. Thus $\hat{b}^*$ and $\hat{V}^*$, the bootstrap estimates of the bias and variance of $\hat{\mu}$, are as given in (3.1) and (3.2). Confidence intervals for μ are as given by (3.3) or (3.4).

8. Simulation Results and Conclusions

A number of simulations were done to evaluate and compare the performances of the jackknife and bootstrap with respect to (a) the estimation of the bias and variance of $\hat{\mu}$ and $\log(\hat{\mu})$ and (b) confidence interval procedures for μ and $\log(\mu)$. For these simulations, it was assumed that X° and Y were independently exponentially distributed with means $\mu = E(X^\circ) = 1$, and $\theta = E(Y) = 4$. These assumptions imply that the censoring rate is 0.20:

$$P(X^\circ > Y) = \int_0^\infty \int_y^\infty \frac{1}{\theta} e^{-y/\theta} \frac{1}{\mu} e^{-x/\mu} \, dxdy = \frac{\mu}{\mu+\theta} = 0.20.$$

Tables 1 and 2 display the simulation results for the bootstrap and jackknife estimators of the bias and standard deviation † of $\hat{\mu}$ and $\log(\hat{\mu})$. The first column of these two tables contains the actual values of the bias and s.d. of $\hat{\mu}$ and $\log(\hat{\mu})$ as estimated by a large simulation of size N = 10,000. The next two columns are, respectively, simulation estimates of the expected values and s.d.'s of the bootstrap and jackknife estimators. The last two columns are n, the original sample size, and N, the number of samples in each simulation. The number of bootstrap samples used was always 500.

Table 3 contains the simulation results for the jackknife, bootstrap, and likelihood confidence interval procedures for μ and $\log(\mu)$. For each confidence interval procedure, Table 3 gives the simulation estimates of the percentage of right and left errors (RE and LE), and the actual coverage rate (percentage of total errors, TE). For the log scale, Table 3 also gives the simulation estimates for the expected confidence interval length of each procedure.

† The standard deviation is, in general, of more interest than the variance, and thus, it was the quantity which was reported.

Table 1: Bias Estimation

Bias $\hat{\mu}$

Jackknife

Actual	$E(\tilde{b})$	$s.d.(\tilde{b})$	n	N
-.0313	-.0307(± .003)	.0977	10	1000
-.0071	-.0074(± .001)	.0225	50	1000

Bootstrap

Actual	$E(\hat{b}^*)$	$s.d.(\hat{b}^*)$	n	N
-.0313	-.0279(± .002)	.0487	10	1000
-.0071	-.0099(± .001)	.0154	50	500

Bias log $(\hat{\mu})$

Jackknife

Actual	$E(\tilde{b}_L)$	$s.d.(\tilde{b}_L)$	n	N
-.0931	-.0892(± .004)	.1149	10	1000
-.0197	-.0204(± .001)	.0261	50	1000

Bootstrap

Actual	$E(\hat{b}_L^*)$	$s.d.(\hat{b}_L^*)$	n	N
-.0931	-.0688(± .002)	.0534	10	1000
-.0197	-.0217(± .001)	.0152	50	500

Table 2: S.D. Estimation

s.d. $(\hat{\mu})$

Jackknife

Actual	$E(\sqrt{\tilde{V}})$	$s.d.\sqrt{\tilde{V}}$	n	N
.3448	.3159(± .006)	.1878	10	1000
.1579	.1555(± .002)	.0490	50	1000

Bootstrap

Actual	$E(\sqrt{\hat{V}^*})$	$s.d.\sqrt{\hat{V}^*}$	n	N
.3448	.2540(± .004)	.1221	10	1000
.1579	.1338(± .002)	.0367	50	500

s.d.(log $(\hat{\mu})$)

Jackknife

Actual	$E(\sqrt{\tilde{V}_L})$	$s.d.(\sqrt{\tilde{V}_L})$	n	N
.3535	.3451(± .005)	.1470	10	1000
.1591	.1597(± .001)	.0396	50	1000

Bootstrap

Actual	$E(\hat{b}_L^*)$	$s.d.(\hat{b}_L^*)$	n	N
.3535	.2907(± .003)	.0947	10	1000
.1591	.1372(± .001)	.0248	50	500

Table 3: 95% Confidence Intervals

For μ

Method	LE	RE	TE	n	N
Jackknife	.146	.005	.151(± .01)	10	1000
Bootstrap	.182	.016	.198(± .01)	10	1000
Bootstrap	.206	.046	.252(± .01)	10	1000
Jackknife	.073	.008	.081(± .01)	50	1000
Bootstrap	.108	.012	.120(± .01)	50	500
Bootstrap	.110	.020	.130(± .01)	50	500

For log (μ)

Method	LE	RE	TE	E(length)	n	N
Jackknife	.089	.019	.108(± .01)	.6764(± .009)	10	1000
Bootstrap	.116	.032	.148(± .01)	.5698(± .006)	10	1000
Bootstrap	.077	.110	.187(± .01)	1.1071(± .005)	10	1000
MLE	.047	.005	.052(± .007)	1.4001(± .004)	10	1000
Likelihood	.033	.028	.061(± .008)	1.4206(± .004)	10	1000
Jackknife	.051	.015	.066(± .01)	.3131(± .003)	50	1000
Bootstrap	.986	.024	.110(± .01)	.2689(± .002)	50	500
Bootstrap	.062	.052	.114(± .01)	.5345(± .001)	50	500
MLE	.018	.016	.034(± .008)	.6204(± .004)	50	1000
Likelihood	.016	.018	.034(± .008)	.6232(± .004)	50	1000

An estimator accurately estimates a parameter only if (a) the expected value of the estimator is close to the actual parameter value and (b) the s.d. of the estimator is small, especially with respect to the expected value of the estimator. In terms of (a), the jackknife estimators of bias and s.d. do quite well. The bootstrap estimators of bias and s.d. don't do quite as well in terms of (a) especially the bootstrap estimators of s.d. which underestimate the actual s.d. In terms of (b) both the jackknife and bootstrap do poorly with respect to bias estimation as the s.d.'s of the estimator are substantially larger than their expected values. However, with respect to (b), both the jackknife and bootstrap do well with respect to s.d. estimation in the sense that the s.d. estimators have s.d.'s substantially smaller than their expected values. Note that in all cases the s.d. of a bootstrap estimator of bias or s.d. is noticeably smaller than the s.d. of the corresponding mafkknife estimator. There seemed to be little advantage in using the log scale instead of the original scale in either bias or s.d. estimation.

A confidence interval procedure performs well only if (a) the actual coverage rate is close to the desired coverage rate and (b) the percentage of right and left errors is roughly equal. In addition, for any two confidence interval procedures with similar (a) and (b) behavior, we prefer the one which yields shorter CI's on the average. The two likelihood CI procedures perform well in terms of (a), but are actually too conservative for $n = 50$. These procedures also perform well in terms of (b) except for the MLE procedure when $n = 10$. The expected CI lengths are, however, long in comparison to the other methods. The jackknife performs

well in terms of (a) when $n = 50$. In terms of (b), its performance is always poor. However, the expected CI length of the jackknife compares favorably to that of any other CI procedures. (Note that expected CI lengths were given only for the log scale.) Both bootstrap CI procedures perform poorly with respect to (a). The percentile method performs well with respect to (b) on the log scale. The expected CI lengths of the percentile method are relatively large, while the other bootstrap CI procedure has expected CI lengths which are relatively short. Overall, it seems that the jackknife is superior to bootstrap as far as CI procedures for $\hat{\mu}$ and $\log(\hat{\mu})$ are concerned.

9. Bibliographic Notes

Basic reviews of jackknife methodology and theory are given by Miller (1974), Efron (1982), Hinkley (1982). The latter two references also discuss bootstrap methods. A specific discussion of bootstrapping with censored data is given by Efrom (1980). A study of influence functions in censored data problems has been published by Reid (1981). General accounts of the Kaplan-Meier estimate may be found in Miller (1980) and in Kalbfleisch & Prentice (1980).

References

Efron, B. (1980). Censored data and the bootstrap. J. Amer. Statist. Assoc., 76, 312-319.

Efron, B. (1982). The Jackknife, the Bootstrap, and Other Resampling Plans. SIAM-SBMS Monograph No. 38. SIAM, Philadelphia.

Hinkley, D.V. (1982). Subsampling Methods: The Jackknife, Bootstrap and Cross-validation. Unpublished monograph.

Kalbfleisch, J.D. & Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.

Miller, R.G. (1974). The jackknife: a review. Biometrika, 61, 1-17.

Miller, R.G. (1981). Survival Analysis. Wiley, New York.

Reid, N.M. (1981). Influence functions for censored data. Ann. Statist., 9, 78-92.