A $J^{th}$-ORDER BOOTSTRAP BIAS-CORRECTED ESTIMATOR
OF THE NUMBER OF CLASSES OF OBJECTS

James L. Norris, III and Duane A. Meeter

Department of Mathematics and Computer Science
Wake Forest University
Winston-Salem, North Carolina 27109

Department of Statistics
Florida State University
Tallahassee, Florida 32306


March 1991

Key words and phrases: Bootstrap Bias-Corrected Estimator, Functional Statistic, Simple Random Sampling, Stratified Random Sampling.
A $j^{th}$-order bootstrap bias-corrected estimator of the number of classes of objects

BY JAMES L. NORRIS, III  
Department of Mathematics and Computer Science, Wake Forest University, Box 7388 Reynolda Station, Winston-Salem, North Carolina 27109, U.S.A.

AND DUANE A. MEETER  
Department of Statistics, The Florida State University, Tallahassee, Florida 32306, U.S.A.

SUMMARY

Under both simple and stratified random sampling, we develop a $j^{th}$-order bootstrap bias-corrected estimator of the number of classes $\theta$ which exist in a study region. Under simple random sampling, we compare this estimator's bias (for $j = 1, 2$) to the bias of the $j^{th}$-order jackknife bias-corrected estimator of $\theta$ developed by Burnham and Overton (1978). The bootstrap estimators' biases can be closer to zero when the classes have sufficiently dense dispersals over the study region. Under stratified random sampling, there are no developed jackknife competitors to our bootstrap estimators; however, estimating the number of classes under this type of sampling can be important for such applications as estimating the (animal) population size in capture-recapture experiments when there are discrete changes in capture probabilities over time.

For both single and multiple populations, we also develop the general form of the $j^{th}$-order bootstrap bias-corrected estimator of a parameter when the initial estimator is its functional statistic.

Some key words: Bootstrap Bias-Corrected Estimator, Functional Statistic, Simple Random Sampling, Stratified Random Sampling
1. INTRODUCTION

Estimating the number of different classes of objects in a region is of interest in many diverse fields. In ecology, the classes might be species on a body of land (Fisher, Corbet, and Williams (1943)) or they might be individual animals on a wildlife preserve (Burnham and Overton (1978)). In linguistics, the classes might be word types in a body of literature (Efron and Thisted (1976)), while in auditing, they might be types of errors in a report (Moors (1983)). Knowledge of the number of classes in a population can be very useful. For example, the number of herbaceous plant species in the region provides a measure of the diversity of the region's herbaceous plants, which is useful in assessing the ecosystem's adaptability to environmental changes; also, the number of grizzly bears in a wildlife preserve provides a criterion to evaluate wildlife preservation programs as well as a tool to assist in the planning of future programs.

In Section 2, we develop a nonparametric estimator (namely a $j^{th}$-order bootstrap bias-corrected estimator) of $\theta$, the number of classes. This extends previous research which has developed bootstrap and $j^{th}$-order jackknife bias-corrected estimators of $\theta$ (see Smith and van Belle (1984) and Burnham and Overton (1978)). We abbreviate the phrase "bootstrap (jackknife) bias-corrected estimator" to "bootstrap (jackknife) estimator". In Section 2, we also develop the general form of a $j^{th}$-order bootstrap estimator of a parameter when the initial estimator is its functional statistic. In Section 3, we extend Section 2's results from a single population to multiple populations. In estimating $\theta$, this extension corresponds to extending simple random sampling results to stratified random sampling. This extension can be very important. For instance, it has applicability towards estimating the number of individual animals (in a preserve) when there are discrete changes in capture probabilities over time. Section 4 presents simulations from data of a large field
study, while Section 5 summarizes this paper.

2. SIMPLE RANDOM SAMPLING

2.1. General form of a \( j \)th-order bootstrap bias-corrected estimator

We first consider the general form of a \( j \)th-order bootstrap estimator of a parameter when the initial estimator is its functional statistic. Let \( \mathcal{F}_0 \) be a probability distribution, and let \( \mu(\mathcal{F}_0) \) be the parameter of interest. Now let \( \mathcal{F}_1 \) be the empirical distribution of \( n \) independent observations from \( \mathcal{F}_0 \), and let the functional statistic \( \mu(\mathcal{F}_1) \) be the initial estimator of \( \mu(\mathcal{F}_0) \). For \( k \geq 1 \), let \( \mathcal{F}_k \) be the empirical distribution of \( n \) independent observations from \( \mathcal{F}_{k-1} \). Define \( \text{Bias}^*(\mathcal{F}_k) = \mathbb{E}_{\mathcal{F}_k}(\mu(\mathcal{F}_{k+1}) - \mu(\mathcal{F}_k)) \). Note that \( \text{Bias}^*(\mathcal{F}_0) = \text{Bias}_{\mathcal{F}_0}(\mu(\mathcal{F}_1)) \), the bias of the initial estimator; also, note that \( \text{Bias}^*(\mathcal{F}_k) = \text{Bias}_{\mathcal{F}_k}(\mu(\mathcal{F}_{k+1})) \).

The initial estimator's bias (denoted \( \text{Bias}^*(\mathcal{F}_0) \), above) is estimated by \( \text{Bias}^*(\mathcal{F}_1) \); then \( \text{Bias}^*(\mathcal{F}_1) \) is subtracted from the initial estimator in order to obtain the bootstrap bias-corrected estimator of \( \mu(\mathcal{F}_0) \) (Efron (1982), pg. 33). We call this estimator the first-order bootstrap (bias-corrected) estimator of \( \mu(\mathcal{F}_0) \); it equals

\[
\mu(\mathcal{F}_1) - \text{Bias}^*(\mathcal{F}_1) = 2 \cdot \mu(\mathcal{F}_1) - \mathbb{E}_{\mathcal{F}_1}(\mu(\mathcal{F}_2)).
\]

(1)

As in Efron's double bootstrap of an error term (Efron (1983)), we "double bootstrap" our bias to obtain a second-order bootstrap estimator. The basic idea is that the first-order bootstrap estimator still has bias so we estimate its (true) bias under \( \mathcal{F}_0 \) by the bias of the corresponding estimator under \( \mathcal{F}_1 \). This estimated bias is subtracted from the
first-order bootstrap estimator in order to obtain the second-order bootstrap estimator. Specifically, \( \text{Bias} \mathcal{F}_0(\mu(\mathcal{F}_1) - \text{Bias}^*(\mathcal{F}_1)) \) is estimated by \( \text{Bias} \mathcal{F}_1(\mu(\mathcal{F}_2) - \text{Bias}^*(\mathcal{F}_2)) \), and the second-order bootstrap estimator is

\[
\mu(\mathcal{F}_1) - \text{Bias}^*(\mathcal{F}_1) - [\text{Bias} \mathcal{F}_1(\mu(\mathcal{F}_2) - \text{Bias}^*(\mathcal{F}_2))].
\] (2)

Now the bias of this estimator (under \( \mathcal{F}_0 \)) can be estimated by the bias of the corresponding estimator under \( \mathcal{F}_1 \). This estimated bias can be subtracted from (2) to obtain the third-order bootstrap estimator. This procedure can be repeated to successively obtain higher-order bootstrap estimators.

**Theorem 2.1.** The \( j^{th} \)-order bootstrap bias-corrected estimator of \( \mu(\mathcal{F}_0) \) is

\[
\sum_{k=1}^{j+1} \binom{j+1}{k} (-1)^{k-1} E \mathcal{F}_1(\mu(\mathcal{F}_k)).
\] (3)

This theorem can be proved in a straightforward manner using induction and Pascal's triangle. (For details, see the Norris and Meeter (1990a) technical report on this subject.)

The \( j^{th} \)-order bootstrap estimator can also be expressed in terms of forward difference operators. We use the following notations:

- \( \Delta \) is a forward difference operator (i.e. \( \Delta[f(y)] = f(y+1) - f(y) \) and
- \( \Delta^n \) is the \( n \)th-convolution of the \( \Delta \) operator
  (e.g. \( \Delta^2[f(y)] = \Delta[\Delta[f(y)]] \) and \( \Delta^0[f(y)] = f(y) \)).

From Johnson and Kotz (1969, pg. 2),
\[ \Delta^n[f(y)] = \sum_{k=0}^{n} \binom{n}{k} (-1)^k f(y+n-k). \]  

By use of (4) and appropriate changing of indices, the \( j \)-th-order bootstrap estimator of \( \mu(\mathcal{F}_0) \) can be expressed as

\[ (-1)^j \left[ \Delta^j \left[ E_{\mathcal{F}_1}(\text{Bias}^*(\mathcal{F}_y)) \right] \bigg|_{y=0} \right] + \mu(\mathcal{F}_0), \]  

where \( \Delta^j \) treats \( E_{\mathcal{F}_1}(\text{Bias}^*(\mathcal{F}_y)) \) as a function of \( y \) only and \( E_{\mathcal{F}_1}(\mu(\mathcal{F}_0)) \) is defined as \( \mu(\mathcal{F}_0) \) (see Norris and Meeter (1990a) for details).

2.2. A \( j \)-th-order bootstrap bias-corrected estimator of \( \theta \)

We now restrict ourselves to the examination of a \( j \)-th-order bootstrap estimator of the number of classes \( \theta \). Let the random variable \( Y_{ij} \) indicate the presence/absence of class \( i \) on sample sector \( j \); in particular, \( Y_{ij} = 1 \) if class \( i \) is present on sample sector \( j \) while \( Y_{ij} = 0 \), otherwise. Let \( \underline{Y}_j \) denote the vector \( (Y_{1j}, Y_{2j}, ..., Y_{\theta j}) \). In Section 2, we assume that \( \underline{Y}_1, \underline{Y}_2, ..., \underline{Y}_m \) are iid observations from a probability distribution \( \mathcal{F}_0 \). In sampling \( m \) quadrats (sectors) from a region in order to estimate the number of herbaceous plant species in that region, the \( \underline{Y}_j \)'s will be iid if we conduct simple random sampling of \( m \) quadrats with replacement from the total quadrats of the region. In capture-recapture studies where we examine \( m \) capture periods (sectors) on a region in order to estimate the number of individual grizzly bears in that region (with \( Y_{ij} = 1 \) iff individual \( i \) is captured on the \( j \)-th capture period), the \( \underline{Y}_j \)'s will be iid if the capture periods are independent and the capture probabilities do not change over time. (However, the chance that individual \( i \) is
Let $\mathcal{F}_k$ denote the empirical distribution of $Y_1, Y_2, \ldots, Y_m$. In general, for $k \geq 1$, let $\mathcal{F}_k$ denote the empirical distribution of $m$ iid observations from $\mathcal{F}_{k-1}$. Furthermore, for $k \geq 0$ we say that class $i$ is "observed" in a particular vector sampled from $\mathcal{F}_k$ if the $i$th component of this vector is 1. We let $\mu(\mathcal{F}_k)$ denote the number of classes for which the class can be observed (i.e. has positive probability of being observed) given that we sample from $\mathcal{F}_k$. Assuming that each class $i$ of the region has some positive probability $p_i$ of being observed on sample sector $j$, $\mu(\mathcal{F}_0) = \theta$. Note that $\mu(\mathcal{F}_1)$ is the number of classes that are observed in (at least one of) the $m$ sample sectors. We also let $S$ denote this number of observed classes. For $k \geq 0$, a particular vector from $\mathcal{F}_k$ may be selected more than once in the formation of $\mathcal{F}_{k+1}$. However, we call each of the $m$ selected vectors from $\mathcal{F}_k$ a "vector of $\mathcal{F}_{k+1}"$. For $k \geq 1$, let $X_{i,k}$ denote the number of vectors of $\mathcal{F}_k$ for which the $i$th component is 1. For brevity, we also express $X_{i,1}$ as $X_i$. Note that $X_i$ is the number of sample sectors that contain class $i$; also, observe that

\[ (for \ k \geq 1) \quad \mu(\mathcal{F}_k) = \sum_{i=1}^{\theta} I(X_{i,k} \geq 1), \] where $I(\cdot)$ denotes the indicator function.

Now from Theorem 2.1, the $j$th bootstrap bias-corrected estimator of $\theta$ is

\[
\sum_{k=1}^{j+1} \binom{j+1}{k} (-1)^{k-1} E_{\mathcal{F}_k}(\mu(\mathcal{F}_k)) = (-1)^j \left\{ \Delta[ E_{\mathcal{F}_1}(\text{Bias}(\mathcal{F}))] \right\}_{y=0} + \mu(\mathcal{F}_0). \tag{1}
\]

Our first-order bootstrap estimator is the same as the bootstrap estimator of Smith and van Belle (1984).

Fortunately, there is a closed-form solution for $E_{\mathcal{F}_1}(\mu(\mathcal{F}_k))$ for each $k \geq 1$ and for each realization of $\mathcal{F}_1$; thus we can obtain our $j$th-order bootstrap estimate without simulation. For $0 \leq a \leq m$ and $b = 0, 1, \ldots, m$, let
\[ t_{(a;b)} = \left( \frac{m}{b} \right)^{a} \left( \frac{a}{m} \right)^{b} \left( 1 - \frac{a}{m} \right)^{(m-b)} , \] (2)

where \( (0)^{0} \) is defined as 1. The stochastic process \( \{X_{i,k}, k = 1,2, \ldots \} \) is a \( m+1 \)-state stationary Markov process, where \( P(X_{i,k+1} = b \mid X_{i,k} = a) = t_{(a;b)} \) for \( a = 0,1,\ldots,m \) and \( b = 0,1,\ldots,m \). Thus, the transition matrix \( P \) of this process has \( t_{(a;b)} \) as its \( (a;b) \) element. (For convenience, the numbering for the rows and columns of matrices begins at zero rather than one.) Let \( P^{(n)} \) denote the matrix which is the \( n \)-th power of \( P \), and let \( \left[ P^{(n)} \right]_{(a;b)} \) denote the \( (a;b) \) element of \( P^{(n)} \). We now have

\[ E_{\mathcal{F}}(\mu(\mathcal{F})) = \sum_{i=1}^{\theta} I(X_{i,k} \geq 1) = \sum_{i=1}^{\theta} P(X_{i,k} \geq 1 \mid X_{1}) = \sum_{i=1}^{\theta} \left[ 1 - \left[ P^{(k-1)} \right]_{(X_{1}',0)} \right] , \] (3)

where the primed index lists the observed classes first. Now we obtain the \( j \)-th-order bootstrap estimator by combining (3) with the left-hand side of (1). We label this estimator as \( \hat{\theta}_{Bj} \).

Let \( N_{l} \) denote the number of classes which occur in exactly \( l \) of the \( m \) sample quadrats, \( l = 0,1,\ldots,m \). (Note that \( N_{l} = \sum_{i=1}^{\theta} I(X_{i}=l) \).) A convenient computational form of \( \hat{\theta}_{Bj} \) is

\[ \hat{\theta}_{Bj} = \sum_{l=1}^{m} \left\{ \sum_{k=1}^{j+1} \left[ \begin{array}{c} j+1 \\ k \end{array} \right] (-1)^{(k-1)} \left[ 1 - \left[ P^{(k-1)} \right]_{(l,0)} \right] \right\} \cdot N_{l} \] (4)

**Example 2.1:** Carothers (1973) examined data on taxi cabs in Edinburg, Scotland, where the number of taxi cabs (i.e. classes) was known to be 420. We utilize his full data set from "sampling scheme A", where the places of observation change with sample times
(i.e. sample sectors). For these data, \( m = 10 \) and the vector \((N_j, j=1,2,\ldots,10) = (142,81,49,7,3,1,0,0,0,0)\). The number of observed classes, \( S \), equals 283. From (4), we have that \( \hat{\theta}_{B1} = 342.6 \) and \( \hat{\theta}_{B2} = 365.2 \).

Example 2.2. Jaccard (1908) examined the distribution of plant species over 52 quadrats in an alpine valley in Switzerland. As discussed in Williams (1964, pg. 79), these quadrats were taken in nine different blocks with the quadrats in a block being adjacent to one another. In an attempt to approximate simple random sampling, we randomly chose one quadrat from each block and used only these nine quadrats as our sample sectors. We obtained \((N_j, j=1,2,\ldots,9) = (23,20,6,9,8,5,3,1,1)\). Thus, \( S = 76; \hat{\theta}_{B1} = 86.3 \), and \( \hat{\theta}_{B2} = 89.5 \).

2.3. Biases of the Bootstrap and Jackknife Estimators

2.3.1. The \( j^{th} \)-order bootstrap's bias

From (1) §2.2, the bias of the \( j^{th} \)-order bootstrap estimator is

\[
\text{Bias}(\hat{\theta}_{Bj}) = E_{\mathcal{F}_0} (\hat{\theta}_{Bj}) - \theta = E_{\mathcal{F}_0} \left[ \sum_{k=1}^{j+1} \binom{j+1}{k} (-1)^{k-1} E_{\mathcal{F}_1} \mu(\mathcal{F}_k) \right] - \theta. \tag{1}
\]

Expectations and probabilities are with respect to \( \mathcal{F}_0 \) unless otherwise specified; thus the \( j^{th} \)-order bootstrap estimator's bias equals
\[
\sum_{i=1}^{\theta} \left\{ E \left[ \sum_{k=1}^{j+1} \left[ (k-1)(-1)^{k-1} I(X_{i,k} \geq 1) \right] - 1 \right] \right\}
\]

\[
= \sum_{i=1}^{\theta} \left\{ \left[ \sum_{k=1}^{j+1} (k-1)(-1)^{k-1} \{ 1 - P(X_{i,k} = 0) \} \right] - 1 \right\}
\]

For \( i = 1,2,\ldots,\theta \), recall that \( p_i \) denotes the probability that class \( i \) is observed on the \( j^{th} \)-sample sector. Let \( t_a \) denote the vector \((t_{(a;0)}, t_{(a;1)}, \ldots, t_{(a;m)})\); then

\[
P(X_{i,k} = 0) = 1 - \left[ t_{m} \cdot p_i \cdot p^{(k-1)} \right]_{(0)},
\]

where \( \left[ \begin{array}{c} v \\ \end{array} \right]_{(0)} \) denotes the initial term in the vector \( v \) and \( P^{(0)} \) is interpreted as the identity matrix. So combining (2) and (3), the \( j^{th} \)-order bootstrap estimator's bias can be expressed as a function of the \( p_i \)'s. Smith and van Belle (1984, pg. 125) have previously shown the bias of the first-order bootstrap estimator.

Note that the bias of the \( j^{th} \)-order bootstrap estimator is of the form \( \sum_{i=1}^{\theta} g_j(p_i) \), where class \( i \) contributes \( g_j(p_i) \) to the bias of the \( j^{th} \)-order bootstrap estimator. Note that the contribution of a class to this bias is either positive, negative, or zero and that the \( j^{th} \)-order bootstrap estimator is unbiased if (and only if) the positive and negative contributions of the \( \theta \) classes cancel. In Section 2.3.3, we present plots of the individual contributions of a class to the biases of the first-order and second-order bootstrap estimators.
2.3.2. The $j^{th}$-order jackknife's bias

Burnham and Overton (1978, eq. 4) show that the (generalized) $j^{th}$-order jackknife estimator of $\theta$ based on the initial estimator $S$ is

$$
\hat{\theta}_{j} = \frac{1}{j!} \sum_{k=0}^{j} (-1)^{k} \left[ j \atop k \right] (m-k)^{j} \left[ m \atop k \right]^{-1} \sum_{1 \leq l_{1} \leq \ldots \leq l_{k}} \hat{\theta}_{m \setminus l_{1}, \ldots, l_{k}},
$$

(1)

where $l_{1}, \ldots, l_{k}$ are $k$ distinct integers from the set $\{1, \ldots, m\}$ and $\hat{\theta}_{m \setminus l_{1}, \ldots, l_{k}}$ refers to the number of classes in the sample after the sample sectors $l_{1}, \ldots, l_{k}$ have been removed. Burnham and Overton (1978, eq. 6) and Smith and van Belle (1984, eq. 4.2) give expressions for the expectation of $\hat{\theta}_{j}$ (and thus the bias of $\hat{\theta}_{j}$). The bias of $\hat{\theta}_{j}$ can be conveniently expressed as the sum of the contributions of each of the $\theta$ classes to the bias. This allows us to produce plots for the $j^{th}$-order jackknife estimator that can be compared to the previously mentioned plots for the $j^{th}$-order bootstrap estimator. From (1),

$$
\text{Bias}(\hat{\theta}_{j}) = E(\hat{\theta}_{j} - \theta)
$$

(2)

$$
= E\left[ \frac{1}{j!} \sum_{k=0}^{j} (-1)^{k} \left[ j \atop k \right] (m-k)^{j} \left[ m \atop k \right]^{-1} \sum_{1 \leq l_{1} \leq \ldots \leq l_{k}} \sum_{i=1}^{\theta} I\{X_{i \setminus l_{1}, \ldots, l_{k}} \geq 1\} \right] - \theta,
$$

where $X_{i \setminus l_{1}, \ldots, l_{k}}$ denotes the number of the remaining sample sectors that contain class $i$ after the $l_{1}, \ldots, l_{k}$ sample sectors have been removed. Upon simplification,
\[ \text{Bias}(\hat{\theta}_j) = \sum_{i=1}^{\theta} \left\{ \frac{1}{j!} \sum_{k=0}^{j} (-1)^{k} \binom{j}{k} (m-k)^i \left[ 1-(1-p_i)^{(m-k)} \right] - 1 \right\}. \]

2.3.3. \textit{Plots of contributions to biases}

For a given \( m \), the contributions of class \( i \) to the biases of \( \hat{\theta}_{B1}, \hat{\theta}_{B2}, \hat{\theta}_{J1}, \) and \( \hat{\theta}_{J2} \) are functions of \( p_i \). (Note that the contribution of class \( i \) to the bias of \( S \) is \( [E(I(X_i \geq 0)) - 1] = -(1-p_i)^m \).) Figures 2.1 and 2.2 are plots of class \( i \)'s contributions for \( m = 10 \) and \( m = 20 \), respectively. Except for a change of scale of the abscissa, plots of class \( i \)'s contributions for \( S, \hat{\theta}_{B1}, \) and \( \hat{\theta}_{J1} \) have been previously shown by Smith and van Belle (1984, their Figures 1 and 2 for \( m = 5 \) and \( m = 10 \), respectively). The main features of (our) Figures 2.1 and 2.2 are described below.

(a) For \( p_i \) exceeding 0.30 when \( m = 10 \) or \( p_i \) exceeding 0.15 when \( m = 20 \), class \( i \)'s contributions are close to zero for all estimators (including \( S \)). Thus classes which are likely to be observed contribute little to the biases of these estimators.

(b) Each of the four bias-corrected estimators may over-correct for the negative bias contribution to \( S \), resulting in positive bias contributions for these estimators. Positive bias contributions are more prominent for the jackknife estimators than for the bootstrap estimators, while the maximum (positive) contribution is higher for \( \hat{\theta}_{J2} \) than for \( \hat{\theta}_{J1} \).

(c) For \( p_i \leq 0.07 \) when \( m = 10 \) and for \( p_i \leq 0.03 \) when \( m = 20 \), the \textit{absolute value of class \( i \)'s contribution to bias} (abbreviated class \( i \)'s absolute contribution) is smallest for \( \hat{\theta}_{J2} \) (in comparison with those of the other four estimators). In other words, for such values of \( p_i \), the bias contribution of class \( i \) is closer to zero for \( \hat{\theta}_{J2} \) than it is for the other four estimators. For slightly larger \( p_i \) (0.07 < \( p_i \leq 0.11 \) when \( m = 10 \) and 0.03 < \( p_i \leq 0.06 \) when \( m = 20 \)), \( \hat{\theta}_{J1} \) shows the smallest absolute contribution of class \( i \). For still larger \( p_i \) (0.11 < \( p_i \leq 0.15 \) when \( m = 10 \) and 0.06 < \( p_i \leq 0.08 \) when \( m = 20 \)), \( \hat{\theta}_{B2} \) has the smallest
Fig. 2.1. Class i's contributions to the biases of estimators: m = 10.
Fig. 2.2. Class $i$'s contributions to the biases of estimators: $m = 20$. 

$p_i$
absolute contribution of class i, while for even larger \( p_1 \) (0.15 < \( p_1 \) ≤ 0.26 when \( m = 10 \) and 0.08 < \( p_1 \) ≤ 0.13 when \( m = 20 \), \( \hat{\theta}_{B_i} \) shows the smallest absolute contribution of class i. Thus, the (first-order and second-order) jackknife estimators' biases are closer to zero when the values of the \( p_1 \)'s are small, while the (first-order and second-order) bootstrap estimators' biases can be closer to zero for more moderate values of the \( p_1 \)'s.

2.4. Variances of the bootstrap and jackknife estimators — under independence

We have previously assumed that \( Y_1, Y_2, \ldots, Y_m \) are independent. In Section 2.4, we additionally assume "independence of the classes" in the sense that we assume \( Y_{1j}, Y_{2j}, \ldots, Y_{\theta_j} \) to be independent (\( j = 1, \ldots, m \)). Under this assumption, we show that the variance of \( \hat{\theta}_{B_j} \) and the variance of \( \hat{\theta}_{J_j} \) can each be expressed in the form \( \theta \sum_{i=1}^{\theta} h_e(p_i) \), where \( h_e(p_i) \) is the contribution of class i to the variance of the respective estimator.

2.4.1. The \( j^{th} \)-order bootstrap's variance

From Section 2.2 equations (1) and (3), and since \( 1 - \left[ \mathbf{p}^{(k-1)} \right]_{(0;0)} = 0 \),

\[
\hat{\theta}_{Bj} = \sum_{i=1}^{\theta} \left[ \sum_{k=1}^{j+1} \binom{j+1}{k}(-1)^{k-1} \left[ 1 - \left[ \mathbf{p}^{(k-1)} \right]_{(X_i;0)} \right] \right] \equiv \sum_{i=1}^{\theta} R_i.
\]

Since the \( Y_{ij} \)'s are independent, the \( R_i \)'s are independent; thus

\[
\text{Var}(\hat{\theta}_{B_j}) = \sum_{i=1}^{\theta} \text{Var}(R_i) = \sum_{i=1}^{\theta} \left\{ \text{E}(R_i^2) - [\text{E}(R_i)]^2 \right\}.
\]
From the derivation of the bias of $\hat{\theta}_{Bj}$, 

$$
E(R_i) = \sum_{k=1}^{j+1} \binom{j+1}{k} (-1)^{(k-1)} \left(1 - \left[ t_m \cdot p_i \cdot p^{(k-1)} \right]_{(0)} \right).
$$

Now by conditioning on the possible values that $X_i$ can take,

$$
E(R_i^2) = 
$$

$$
\sum_{k=1}^{j+1} \sum_{n=1}^{j+1} \binom{j+1}{k} \binom{j+1}{n} (-1)^{(n+k)} \sum_{x=0}^{m} \left\{ \left[ 1 - \left[ p^{(k-1)} \right]_{(x,0)} \right] \left[ 1 - \left[ p^{(n-1)} \right]_{(x,0)} \right] t_{(m p_i; x)} \right\}.
$$

Therefore (under independence of the $Y_{ij}$'s), the contribution of Class $i$ to the variance of $\hat{\theta}_{Bj}$ can be found by using (1), (2), and (3).

2.4.2. The $j^{th}$-order jackknife's variance

From equation (2) of Section 2.3.2,

$$
\hat{\theta}_{jj} = \sum_{i=1}^{\theta} \left[ \frac{1}{j!} \sum_{k=0}^{j} (-1)^k \binom{j}{k} \binom{m}{k} (m-k)^{j-1} \sum_{l_1, \ldots, l_k} I[X_{i l_1, \ldots, l_k} \geq 1] \right] = \sum_{i=1}^{\theta} D_i.
$$

Since the $Y_{ij}$'s are independent, the $D_i$'s are independent; thus

$$
\text{Var}(\hat{\theta}_{jj}) = \sum_{i=1}^{\theta} \text{Var}(D_i) = \sum_{i=1}^{\theta} \left\{ E(D_i^2) - [E(D_i)]^2 \right\}.
$$
From the derivation of the bias of $\hat{\theta}_{jj}$ (see Section 2.3.2)

$$E(D_i) = \frac{1}{j!} \sum_{k=0}^{j} (-1)^k C_j^k \left( \frac{m-k}{k} \right)^j (m-k)^j \left( 1 - (1-p_i)^{(m-k)} \right).$$ (3)

To derive $E(D_i^2)$, first note that there are $\binom{m}{k} \cdot \binom{m}{n}$ possible pairings of subsets of size $k$ (from $m$ elements) with subsets of size $n$ (from the same $m$ elements). The number of these pairs which have $w$ common elements is $\binom{m}{k} \cdot \binom{k}{w} \cdot \binom{m-k}{n-w}$, where this product is set to zero if undefined (i.e. if $w > k$, $m-k < n-w$, or $n-w < 0$). Also, note that if $\{l_1, l_2, \ldots, l_k\}$ is one subset of $k$ sectors from the $m$ sample sectors and $\{d_1, d_2, \ldots, d_n\}$ is another such subset and if these two subsets have $w$ sectors in common then

$$E\left[I(X_{l_1}, \ldots, l_k \geq 1) I(X_{d_1}, \ldots, d_n \geq 1)\right] = \left[ 1 - (1-p_i)^{(m-k)} \right] \cdot \left[ (1-p_i)^{(m-n)} + (1-p_i)^{(m-w)} \right],$$ (4)

which, for convenience, is defined to be $f_{iknw}$.

Therefore from (1) and (4),

$$E(D_i^2) = \frac{1}{j!} \sum_{k=0}^{j} \sum_{n=0}^{k} \sum_{w=0}^{k-n} (-1)^{k+n} \binom{j}{k} \binom{j}{n} \binom{m-k}{n} \binom{m-n}{w} \binom{m-k}{n-w} \cdot f_{iknw},$$ (5)

where $\binom{m-k}{n-w}$ is set to zero when it is undefined. Therefore the contribution of class $i$ to the variance of $\hat{\theta}_{jj}$ can be found by using (2), (3), and (5).
2.4.3. Plots of contributions to variances

The contributions of class i to the variances of $\hat{S}$, $\hat{\theta}_{B_1}$, $\hat{\theta}_{B_2}$, $\hat{\theta}_{J_1}$, and $\hat{\theta}_{J_2}$ are functions of $p_i$. Figures 2.3 and 2.4 are plots of class i’s contributions for $m = 10$ and $m = 20$, respectively. For either of these two sampling intensities, if $p_i \geq 0.5$ then the contributions to variance are small for all estimators. For either sampling intensity and under independence, \( \text{Var}(\hat{\theta}_{J_2}) > \text{Var}(\hat{\theta}_{J_1}) > \text{Var}(\hat{\theta}_{B_2}) > \text{Var}(\hat{\theta}_{B_1}) > \text{Var}(S) \) (as long as $p_i \leq 0.5 \ \forall \ i$) with the variance of $\hat{\theta}_{J_2}$ being at least twice as large as the others.

2.5. Mean-Square Error Computations

If we have predictions concerning $\theta$ and the $p_i$'s then the bias and variance contributions can assist us in predicting the performances of the estimators. It is useful to define the relative bias (variance) of an estimator of $\theta$ as the bias (variance) of the estimator divided by $\theta$. Under independence of the $Y_{ij}$'s, the relative biases (variances) of $S$, $\hat{\theta}_{B_1}$, $\hat{\theta}_{B_2}$, $\hat{\theta}_{J_1}$, and $\hat{\theta}_{J_2}$ are respective averages of individual class contributions.

Example 2.3: Suppose we predict that the fraction of the $p_i$'s which take a particular value $p$ is proportional to $p(1-p)^9$ for $p = .01, .02, \ldots, .40$ (a discretized version of a beta(2,10) distribution). Further (for variance contributions) assume independence of the $Y_{ij}$’s. Then for $m = 20$ and sufficiently large $\theta$, the relative biases (variances) of $S$, $\hat{\theta}_{B_1}$, $\hat{\theta}_{B_2}$, $\hat{\theta}_{J_1}$, and $\hat{\theta}_{J_2}$ are $-0.121 (.077)$, $-0.032 (.123)$, $-0.017 (.158)$, $0.039 (.257)$, and $0.044 (.807)$, respectively. Since the mean-square error of an estimator equals its variance plus the square of its bias, if we additionally have $\theta = 100$ then the mean-square errors of the above five estimators are approximately 153.8, 22.7, 18.7, 40.7, and 100.2, respectively. A FORTRAN program is available from the authors to examine relative biases, relative
Fig. 2-3. Class i's contributions to the variances of estimators: m = 10.
Fig. 2.4. Class i's contributions to the variances of estimators: \( m = 20 \).
variances, and mean square errors for other sets of $p_i$'s.

Assuming independence of the $Y_{ij}$'s, the mean-square error of each of the above estimators, $S$, $\hat{\theta}_{B1}$, $\hat{\theta}_{B2}$, $\hat{\theta}_{J1}$, and $\hat{\theta}_{J2}$, can be expressed in the form $\theta \cdot \left[ \theta \cdot (\bar{b})^2 - \bar{v} \right]$, where $\bar{b}$ denotes the estimator's relative bias and $\bar{v}$ denotes the estimator's relative variance. Therefore when $\theta$ is large, it is particularly useful to reduce $|\bar{b}|$ even if it results in moderate increases to $\bar{v}$. The bias-corrected estimators tend to reduce $|\bar{b}|$ while increasing $\bar{v}$.

3. Stratified Random Sampling

Stratified random sampling of sectors can be useful to practitioners. For instance, biologists often obtain stratified random quadrats from the field in order to assure some samples in each environmental type as well as to reduce the variability of many of their estimators. In regards to examining classes of objects (e.g. species of herbaceous plants) Norris and Meeter (1990b) showed that the expected number of observed classes can be higher under stratified random sampling of quadrats than under simple random sampling of quadrats. In capture/recapture studies for estimating the number of individuals (say, number of grizzly bears in a wildlife refuge), using the concept of stratified random sampling allows us to relax the assumption that capture probabilities do not change over capture periods to the assumption that these probabilities need only be the same within each group (stratum) of capture periods (and not between strata). The different strata may correspond to many things (e.g. different seasons of the year, different parts of the day, or different sampling intensities). In addition, there may be some classes which have extremely small chances of being observed in one stratum (and thus each of these would
essentially contribute a full -1 to the biases of \( S \) and the bias-corrected estimators if only that stratum is sampled) but these same classes may have much larger chances of being observed in another stratum.

We first extend (to a finite number of populations) the results of Section 2.1 on bootstrap bias-correction for general functional statistics. We then use these results to obtain a \( j^{th} \) order bootstrap bias-corrected estimator of \( \theta \) under stratified random sampling of sectors.

Let \( \mathcal{F}_{0}, \mathcal{F}_{1}, ..., \mathcal{F}_{u,0} \) be \( u \) probability distributions, and let \( \bar{\mu}(\mathcal{F}_{0},...,\mathcal{F}_{u,0}) = \bar{\mu}_{0} \) be the parameter of interest. For \( k \geq 1 \) and \( \ell = 1, ..., u \), let \( \mathcal{F}_{\ell,k} \) denote the empirical distribution of \( n_{\ell} \) iid observations from \( \mathcal{F}_{\ell,k-1} \). For convenience, abbreviate the \( u \)-tuple \( (\mathcal{F}_{1,k}, ..., \mathcal{F}_{u,k}) \) as \( \mathcal{F}_{k} \) and define \( \text{Bias}(\mathcal{F}_{k}) \) as \( E_{\mathcal{F}_{k}}(\bar{\mu}(\mathcal{F}_{k+1}) - \bar{\mu}(\mathcal{F}_{k})) \). With the functional statistic \( \bar{\mu}(\mathcal{F}_{1}) \) as the initial estimator of \( \bar{\mu}_{0} \), we have (analogously to Theorem 2.1) that the \( j^{th} \) order bootstrap estimator of \( \bar{\mu}_{0} \) is

\[
\sum_{k=1}^{j+1} \binom{j+1}{k} (-1)^{k-1} E_{\mathcal{F}_{1}}(\bar{\mu}(\mathcal{F}_{k})) = (-1)^{j} \left[ \Delta_{1}[E_{\mathcal{F}_{1}}(\text{Bias}(\mathcal{F}_{j}))] \right]_{y=0} + \bar{\mu}(\mathcal{F}_{0}).
\]

We now restrict ourselves to the examination of a \( j^{th} \) order bootstrap estimator of \( \theta \) under stratified random sampling of sectors. We consider that there are \( u \) strata from which we take \( m_{\ell} \) of the \( m \) sample sectors from the \( \ell^{th} \) stratum. The random variable \( Y_{i,\ell} = 1 \) if class \( i \) is present on the \( j^{th} \) sample sector of stratum \( \ell \), while \( Y_{i,\ell} = 0 \), otherwise. Also, \( \underline{Y}_{\ell} \) denotes the vector \( (Y_{1,\ell}, ..., Y_{\theta_{\ell}}) \). In Section 3, we assume that \( Y_{\ell_{1}}, ..., Y_{\ell_{m_{\ell}}} \) are iid observations from a probability distribution \( \mathcal{F}_{\ell_{0}} \), but we only assume independence and not identical distributions for \( \underline{Y}_{\ell} \) vectors of different strata.

Let \( \mathcal{F}_{\ell,1} \) denote the empirical distribution of \( Y_{\ell_{1}}, Y_{\ell_{2}}, ..., Y_{\ell_{m_{\ell}}} \). For \( k \geq 1 \), let
\( \mathcal{F}_{L_k} \) denote the empirical distribution of \( m_\ell \) iid observations (vectors) from \( \mathcal{F}_{L_{k-1}} \); we call each of these vectors a "vector of \( \mathcal{F}_{L_k} \)". (Recall that we abbreviate the u-tuple \((\mathcal{F}_{1,k}, \ldots, \mathcal{F}_{u,k})\) by \( \mathcal{F}_{k} \).) Similarly to Section 2, \( \hat{\mu}(\mathcal{F}_{k}) \) denotes the number of classes for which the class could be observed if we sample from \( \mathcal{F}_{k} \). Assuming that each class of the region has some positive probability of being observed on a sample sector of some stratum and assuming that \( m_\ell \geq 1 \) \( \forall \ell \), \( \hat{\mu}(\mathcal{F}_{0}) = \theta \). Analogously to Section 2, we let \( \hat{\mu}(\mathcal{F}_{1}) = \tilde{S} \), the number of classes that are observed in (at least one of) the sample sectors.

The \( j \)th-order bootstrap bias-corrected estimator of \( \theta \) is

\[
\hat{\theta}_{Bj} = \sum_{k=1}^{j+1} \binom{j+1}{k} (-1)^{(k-1)} E_{\mathcal{F}_{1}}(\hat{\mu}(\mathcal{F}_{k})) = (-1)^{j} \left[ \Delta^{j} E_{\mathcal{F}_{1}}(\hat{\text{Bias}}(\mathcal{F}_{y})) \right] \bigg|_{y=0} + \hat{\mu}(\mathcal{F}_{0}).
\]

There is a closed-form solution for \( E_{\mathcal{F}_{1}}(\hat{\mu}(\mathcal{F}_{k})) \) for each \( k \geq 1 \) and for each realization of \( \mathcal{F}_{1} \); therefore we do not need simulation in order to obtain the \( j \)th-order bootstrap estimator of \( \theta \).

We now develop these closed-form solutions. For \( i = 1, \ldots, g \) and \( k = 1, 2, \ldots \), we define the \( u \)-dimensional random vector \( \tilde{X}_{i,k} \) as follows: its \( \ell \)th dimension is the number of vectors of \( \mathcal{F}_{L_k} \) for which the \( i \)th component is 1. Let the vector \((a_1, \ldots, a_u)\), which we denote by \( a \), and the vector \((b_1, \ldots, b_u)\), which we denote by \( b \), have the property that \( 0 \leq a_\ell \leq m_\ell \) and \( b_\ell \in \{0,1,\ldots, m_\ell\} \) for each \( \ell = 1, \ldots, u \). Then we set

\[
\tilde{t}_{(a;b)} = \prod_{\ell=1}^{u} \left[ \frac{m_\ell}{b_\ell} \right]^{a_\ell} \left[ \frac{a_\ell}{m_\ell} \right]^{b_\ell} \left[ 1 - \frac{a_\ell}{m_\ell} \right]^{(m_\ell - b_\ell)},
\]

where recall \( 0^0 \) is defined as \( 1 \). The process \( \{\tilde{X}_{i,k}: k = 1, 2, \ldots\} \) is a \( \prod_{\ell=1}^{u} (m_{\ell+1}) \)-state Markov chain. Note that
\[
P\left[ \tilde{X}_{i,k+1} = b \mid \tilde{X}_{i,k} = a \right] = \tilde{t}_{(a;b)}
\]

for \( a_\ell = 0,1,...,m_\ell \) and \( b_\ell = 0,1,...,m_\ell \) \((\ell = 1,...,u)\). Thus the transition matrix \( \tilde{P} \) of this process has \( \tilde{t}_{(a;b)} \) as its \((a;b)\) element, where we use a new and obvious convention to label the rows and columns of matrices. Let \( \mathbf{0} \) represent a \( u \)-dimensional vector of \( 0 \)'s. For \( k \geq 1 \),

\[
E_{\mathcal{F}_k} (\tilde{P}(\mathcal{F}_k)) = \sum_{i=1}^{\theta} \left[ 1 - \left[ \tilde{P}^{(k-1)} \right]_{(\tilde{X}_{i,1};0)} \right] = \sum_{i=1}^{\tilde{S}} \left[ 1 - \left[ \tilde{P}^{(k-1)} \right]_{(\tilde{X}_{i,1};0)} \right],
\]

where \( \left[ \tilde{P}^{(n)} \right]_{(a;b)} \) is the \((a;b)\) element in the matrix which is the \( n \)\textsuperscript{th} power of \( \tilde{P} \). Now

\[
\tilde{b}_{Bj} = \sum_{i=0}^{\theta} \left\{ \sum_{k=1}^{j+1} \left( \begin{array}{c} j+1 \\ k \end{array} \right) (-1)^{k-1} \left[ 1 - \left[ \tilde{P}^{(k-1)} \right]_{(\tilde{X}_{i,1};0)} \right] \right\} = \sum_{i=0}^{\theta} \tilde{R}_i. \quad (2)
\]

This form of \( \tilde{b}_{Bj} \) motivates discussion of the bias and variance of this estimator.

Let \( m_{p_i} \) denote the \( u \)-dimensional vector whose \( \ell \)\textsuperscript{th} component is \( m_{\ell}' p_{i \ell} \), and let \( \tilde{r}_{m_{p_i}} \) denote the \( \prod_{\ell=1}^{u} (m_{\ell}+1) \) dimensional vector whose \((b_1,...,b_u)\) element is \( \tilde{r}_{(m_{p_i};b)} \). Then

\[
\text{Bias}(\tilde{b}_{Bj}) = \sum_{i=0}^{\theta} \left\{ \left[ \sum_{k=1}^{j+1} \left( \begin{array}{c} j+1 \\ k \end{array} \right) (-1)^{k-1} \left[ 1 - \left[ \tilde{r}_{m_{p_i}} \cdot \tilde{P}^{(k-1)} \right]_{(0)} \right] \right] - 1 \right\}, \quad (3)
\]

where \( (\mathbf{v})_{(0)} \) denotes the \( 0 \) element of \( \mathbf{v} \). Note that the bias of \( \tilde{b}_{Bj} \) is of the form
\[ \sum_{i=1}^{\theta} \tilde{g}_j(p_{i1},...,p_{iu}), \text{ where class } i \text{ contributes } \tilde{g}_j(p_{i1},...,p_{iu}) \text{ to the bias of } \tilde{\theta}_{ Bj}. \]

If we make the additional assumption that the \( Y_{ij} \)'s are independent then \( \text{Var}(\tilde{\theta}_{ Bj}) \) can also be expressed as a sum of individual classes' contributions. Assuming the above independence and using (2),

\[
\text{Var}(\tilde{\theta}_{ Bj}) = \sum_{i=0}^{\theta} \left\{ E(\tilde{R}_i^2) - \left[ E(\tilde{R}_i) \right]^2 \right\}.
\]

By first conditioning on \( X_{i,1} \),

\[
E(\tilde{R}_i^2) = \sum_{b=0}^{m} \sum_{k=1}^{j+1} \sum_{n=1}^{j+1} \left[ \begin{array}{c} j+1 \\ k \end{array} \right] \left[ \begin{array}{c} j+1 \\ n \end{array} \right] (k+n) \left[ 1 - \left( P_{(k)}(b;0) \right) \right] \left[ 1 - \left( P_{(n-1)}(b;0) \right) \right] \tilde{t}(mp_i;b),
\]

where the first summation symbol refers to summing over all \( \prod_{\ell=1}^{u} (m_{\ell}+1) \) possible values for \( b \).

The components \( \tilde{P} \) and \( \tilde{t} \) can be related to their srs counterparts, \( P \) and \( t \), by

\[
\left[ \tilde{P}(k) \right]_{(b;0)} = \prod_{\ell=1}^{u} \left[ P(k) \right]_{(b;\ell;0)} \quad \text{and} \quad \tilde{t}(mp_i;b) = \prod_{\ell=1}^{u} t(m_{\ell},p_{i\ell};b;\ell).
\]

Also note that \( E(\tilde{R}_i) \) is the term inside the braces of (3); therefore \( \text{Var}(\tilde{\theta}_{ Bj}) \) can be expressed in the form

\[
\sum_{i=1}^{\theta} \tilde{h}_j(p_{i1},...,p_{iu}) \text{ where class } i \text{ contributes } \tilde{h}_j(p_{i1},...,p_{iu}) \text{ to the variance of } \tilde{\theta}_{ Bj}. \]

For \( u = 2 \) (i.e. two strata) and \( m_1 = m_2 = 10 \), we now examine plots of class \( i \)'s contributions to the biases and variances of \( \tilde{S}, \tilde{\theta}_{B1}, \) and \( \tilde{\theta}_{B2} \). These contributions are functions of \( p_{i1} \) and \( p_{i2} \). Figures 3.1, 3.2, and 3.3 present contour plots of class \( i \)'s contributions to the biases of \( \tilde{S}, \tilde{\theta}_{B1}, \) and \( \tilde{\theta}_{B2} \), respectively, while Figures 3.4, 3.5,
and 3.6 present contour plots of class i's contributions to the variances of the respective estimators. On each of the contour plots, the values of the contours are shown on the first line below the horizontal axis.

Note the symmetry of each of the contours about the line $p_{12} = p_{11}$. Also observe that each contour can be approximated by a line of the form $p_{11} + p_{12} = a_c$, where $a_c$ depends on the contour. Analogously to srs, as $p_{11}$ and $p_{12}$ increase from the origin, $\bar{\theta}_{B2}$'s bias contribution approaches zero faster than does that of $\bar{\theta}_{B1}$ or $\bar{S}$; however, $\bar{\theta}_{B2}$ may overcorrect resulting in positive bias contributions. For all $(p_{11}, p_{12})$ pairs, $\bar{\theta}_{B2}$'s variance contribution is greater than that of $\bar{\theta}_{B1}$ which is greater than that of $S$. For each of the three estimators, the variance contribution is maximized when $p_{11} + p_{12} = 0.06$.

If we have predictions concerning $\theta$ and the $p_{i\ell}$'s then the bias and variance contributions can assist us in predicting the performances of the estimators.

**Example 3.1:** As a simple illustration, suppose that the fraction of the $(p_{i1}, p_{i2})$'s which take a particular value $(p_1, p_2)$ is proportional to $1-p_1-p_2$ for $p_1 = .05, .1, .15, .2, .25$ and $p_2 = .05, .1, .15, .2, .25$. Further (for variance contributions) assume independence of the $Y_{i\ell}$'s; then for $m_1 = m_2 = 10$ and for sufficiently large $\theta$, the relative biases (variances) of $S$, $\bar{\theta}_{B1}$, and $\bar{\theta}_{B2}$ are $-0.078 (.064)$, $0.015 (.099)$, and $0.027 (.128)$, respectively. If we additionally have $\theta = 100$ then the mean-square errors of the above three estimators are 67.8, 12.1, and 20.0, respectively.

4. **Simulations on a field study**

Professor Robert Livingston of The Florida State University graciously allowed us to examine bias-corrected estimators on his spatial distribution survey data. Details of his study are discussed in Livingston (1987). His data consist of 900 core samples of infaunal
Fig. 3.1. Class i's contribution to the bias of the number of observed classes: $m_1 = m_2 = 10$. 
Fig. 3.2. Class i’s contribution to the bias of the first-order bootstrap: $m_1 = m_2 = 10$. 
Fig. 3.3. Class i’s contribution to the bias of the second-order bootstrap: \( m_1 = m_2 = 10 \).
Fig. 3·4. Class i's contribution to the variance of the number of observed classes: $m_1 = m_2 = 10$. 
Fig. 3·5. Class i’s contribution to the variance of the first-order bootstrap: \( m_1 = m_2 = 10 \).
Fig. 3.6. Class $i$'s contribution to the variance of the second-order bootstrap: $m_1 = m_2 = 10$. 
macroinvertebrates. Three sampling locations in Apalachicola Bay and three sampling times (A, B, and C) per location cause a natural partitioning of the 900 cores into 9 divisions of 100 cores per division.

For each location separately, we consider the 200 cores of times A and B as the (total) region with time A cores comprising stratum 1 and time B cores comprising stratum 2. On each of 5000 replications (per location), we obtain both a simple random sample of \( m = 20 \) sectors (i.e. cores) from the (200 cores of the) region as well as a stratified random sample of \( m_1 = m_2 = 10 \) cores. For each replication, \( S, \hat{\theta}_{B1}, \hat{\theta}_{B2}, \hat{\theta}_{J1}, \hat{\theta}_{J2}, \bar{S}, {\bar{\theta}}_{B1}, \) and \( {\bar{\theta}}_{B2} \) are computed. Sampling was conducted without replacement even though the estimators were theoretically developed under sampling with replacement.

Recall from earlier Figures 2.1 and 2.2 that the jackknife estimators, \( \hat{\theta}_{J1} \) and \( \hat{\theta}_{J2} \), are better in correcting for highly negative relative biases of \( S \) (associated with small \( p_i \)'s) than are \( \hat{\theta}_{B1} \) and \( \hat{\theta}_{B2} \). For each of the locations, the relative bias of \( S \) was highly negative (the least negative being \( -0.37 \)) and the (simulated) biases and mean-square errors of \( \hat{\theta}_{J1} \) and \( \hat{\theta}_{J2} \) were at least 10\% smaller in absolute value than the respective components of \( \hat{\theta}_{B1} \) and \( \hat{\theta}_{B2} \), despite the (at least 50\%) higher variances of the jackknife estimators. The stratified random sampling estimators, \( \bar{S}, {\bar{\theta}}_{B1}, \) and \( {\bar{\theta}}_{B2} \), had biases, mean-square errors, and variances that were extremely close to (within 3\% of) the respective terms of their simple random sampling counterparts.

5. CONCLUSIONS

Under both simple and stratified random sampling, we developed a \( j^{th} \)-order bootstrap bias-corrected estimator of \( \theta \). Under srs, we compared it (for \( j = 1, 2 \) and \( m = 10, 20 \)) to that of Burnham and Overton's \( j^{th} \)-order jackknife bias-corrected estimator of \( \theta \). Differences between the biases of these estimators depend on \( m \) and the \( p_i \)'s with
the jackknife estimators' biases being closer to zero when the values of the $p_i$'s are small, while the bootstrap estimators' biases can be closer to zero for more moderate $p_i$'s. The variances of the jackknife estimators were consistently larger than those of the bootstrap estimators. The $j$th-order bootstrap estimators for stratified random samples can be important for such applications as estimating the population size in capture-recapture experiments when there are discrete changes in capture probabilities over time.

For both single and multiple populations, we also developed the general form of the $j$th-order bootstrap bias-corrected estimator of a parameter when the parameter's functional statistic is the initial estimator.

REFERENCES


