COMPARISON OF RATIO ESTIMATORS BASED ON INTERPENETRATING SUBSAMPLES WITH OR WITHOUT JACKKNIFING

BY
SUBIR GHOSH AND ROBERTO GOMEZ

TECHNICAL REPORT NO. 95
AUGUST 1984

PREPARED UNDER THE AUSPICES OF
PUBLIC HEALTH SERVICE GRANT 5 RO1 GM21215-10

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
COMPARISON OF RATIO ESTIMATORS BASED ON INTERPENETRATING SUBSAMPLES WITH OR WITHOUT JACKKNIFING

By

Subir Ghosh and Roberto Gomez

TECHNICAL REPORT NO. 95
August 1984

PREPARED UNDER THE AUSPICES OF
PUBLIC HEALTH SERVICE GRANT 5 RO1 GM21215-10

DIVISION OF BIOSTATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Comparison of Ratio Estimators Based on Interpenetrating Subsamples With or Without Jackknifing

by

Subir Ghosh

and

Roberto Gomez

Department of Statistics
University of California, Riverside

SUMMARY

In this paper a comparison of four ratio estimators based on interpenetrating subsamples and with or without jackknifing is done both theoretically and empirically with respect to bias, variance and mean square error.

Short Running Title: Comparison of Ratio Estimators

Key words and phrases: Bias, Interpenetrating subsampling, Jackknife, Mean Square error, Ratio estimation, Simple random sample, Variance.
1. **Introduction**

The method of interpenetrating subsamples (IPS) in large scale sampling as introduced in Mahalanobis (1944, 1946), is to draw a sample in the form of two or more subsamples under the same probability sampling design so that each subsample provides a valid estimate of the parameter of interest. The purpose of IPS is to assess both sampling and nonsampling errors in the estimation of the parameter of interest. The United Nations Subcommission on statistical sampling (1949) has recommended the use of IPS with a suggestion of an alternative name "Replicated Sampling". The method of ratio estimation is common in large scale sample surveys in estimating various ratios and the ratio estimator is biased. The Quenouille-Tukey jackknife (see Miller (1974)) gives nonparametric estimators of bias and variance.

Consider a population of $N$ units with $y$ as the variable of interest and $x$ as an auxiliary variable. Denote the population totals of the variables $x$ and $y$ over $N$ population units by $X$ and $Y$. The population ratio $R = (Y/X)$ is an unknown parameter of interest. We shall draw inference on $R$ on the basis of $k$ interpenetrating subsamples of size $m$ each, $(x_{ij}, y_{ij}) i = 1, \ldots, m, j = 1, \ldots, k$. We consider four interpenetrating estimators of $R$ for any probability sampling design, two of which are jackknife interpenetrating subsample estimators (JIPS) and the other two are just interpenetrating subsample (IPS) estimators. We compare these four estimators theoretically and also empirically with respect to bias, variance and mean square error (MSE). In Sections 3 and 4, we do this comparison when the sampling design is simple random sample with replacement.

Let $\hat{Y}_j$ and $\hat{X}_j$ be unbiased estimators of $Y$ and $X$ from the $j$th interpenetrating subsample $(j = 1, \ldots, k)$. Denote $\hat{Y} = (\sum_{j=1}^{k} \hat{Y}_j/k)$ and $\hat{X} = (\sum_{j=1}^{k} \hat{X}_j/k)$. 
Two IPS estimators of \( R \) are given by

\[
\hat{R}_1 = \frac{\hat{Y}}{\hat{X}} \quad \text{and} \quad \hat{R}_2 = \frac{1}{k} \sum_{j=1}^{k} \frac{\hat{Y}_j}{\hat{X}_j}.
\]

Denote \( \hat{Y}_j(u) = \sum_{j=1, j \neq u}^{k} \hat{Y}_j/(k-1) \), \( \hat{X}_j(u) = \sum_{j=1, j \neq u}^{k} \hat{X}_j/(k-1) \), \( \hat{R}_1(u) = (\hat{Y}_j(u)/\hat{X}_j(u)) \) and \( \hat{R}_1(*) = \frac{1}{m} \sum_{v=1}^{m} \hat{R}_1(u)/k \). The first JIPS estimator is then

\[
\hat{R}_3 = k\hat{R}_1 - (k-1)\hat{R}_1(*).
\]

Let \( \hat{Y}_j(v) \) and \( \hat{X}_j(v) \) be unbiased estimators of \( Y \) and \( X \) from the \( j \)th inter-penetrating subsample eliminating the \( v \)th unit. Denote \( r_j = \hat{Y}_j/\hat{X}_j \),

\[
r_j(v) = \frac{\hat{Y}_j(v)}{\hat{X}_j(v)}, \quad r_j(*) = \frac{1}{m} \sum_{v=1}^{m} r_j(v), \quad \hat{R}_2(v) = \frac{1}{k} \sum_{j=1}^{k} r_j(v) \quad \text{and} \quad \hat{R}_2(*) = \frac{1}{m} \sum_{v=1}^{m} \hat{R}_2(v).
\]

The second JIPS estimator is then

\[
\hat{R}_4 = \frac{1}{k} \sum_{j=1}^{k} [mr_j - (m-1)r_j(*)] = m\hat{R}_2 - (m-1)\hat{R}_2(*).
\]

Note that for \( k = 2 \), \( \hat{R}_2 = \hat{R}_1(*) \) and hence from (2),

\[
\hat{R}_1 = (\hat{R}_2 + \hat{R}_3)/2.
\]

For \( k = 2 \), from theoretical comparison, we find \( \hat{R}_1 \) and \( \hat{R}_3 \) are equally better over \( \hat{R}_2 \) and \( \hat{R}_4 \). For \( k > 2 \), both theoretical and empirical comparisons suggest the superiority of \( \hat{R}_1 \).

The purpose of this study is to investigate whether jackknifing has any positive impact on ratio estimation based on interpenetrating subsamples, in terms of reducing bias and mean square error.
2. Bias, Variance and MSE

Denote the bias of $\hat{R}_i$ to the second degree approximation by $B_i$ ($i=1,2,3,4$), the variance of $\hat{R}_i$ to the second degree approximation by $V_i$ and

\[
B_3^{(1)} = \frac{1}{k} \sum_{u=1}^{k} \{ R \text{Var}(\hat{X}_u) - \text{Cov}(\hat{X}_u, \hat{Y}_u) \},
\]

\[
B_3^{(2)} = R \text{Var}(\hat{X}) - \text{Cov}(\hat{X}, \hat{Y}),
\]

\[
B_4^{(1)} = \frac{1}{m} \sum_{j=1}^{k} \sum_{v=1}^{m} \{ R \text{Var}(\hat{X}_{j(v)}) - \text{Cov}(\hat{X}_{j(v)}, \hat{Y}_{j(v)}) \},
\]

\[
B_4^{(2)} = \sum_{j=1}^{k} \{ R \text{Var}(\hat{X}_j) - \text{Cov}(\hat{X}_j, \hat{Y}_j) \}.
\]

It can be seen in Murthy (1967) that

\[
B_1 = \frac{B_4^{(2)}}{k^2 \chi^2} \quad \text{and} \quad B_2 = kB_1,
\]

The estimators of $B_i$, $i = 1, 2, 3, 4$, are as follows

\[
\hat{B}_1 = \frac{1}{k-1} (\hat{R}_2-\hat{R}_1),
\]

\[
\hat{B}_2 = k \hat{B}_1,
\]

\[
\hat{B}_3 = (k-1)(\hat{R}_1(\cdot)-\hat{R}_1),
\]

\[
\hat{B}_4 = (m-1)(\hat{R}_2(\cdot)-\hat{R}_2).
\]

We now take the expectations of $\hat{B}_3$ and $\hat{B}_4$ assuming $\hat{R}_i(\cdot)$ and $\hat{R}_i$, $i = 1, 2$, as estimators of $R$.

\[
E(\hat{B}_3) = \frac{k-1}{\chi^2} [B_3^{(1)} - B_3^{(2)}],
\]

\[
E(\hat{B}_4) = \frac{m-1}{k\chi^2} [B_4^{(1)} - B_4^{(2)}].
\]
Assume now

\[(9) \quad \text{Cov}(\hat{X}_j, \hat{X}_{j'}) = \text{Cov}(\hat{Y}_j, \hat{Y}_{j'}) = \text{Cov}(\hat{X}_j, \hat{Y}_{j'}) = 0 \quad \text{for} \quad j \neq j'. \]

The assumption (9) means that the estimators based on two different subsamples are uncorrelated.

**Lemma 1.** Under (9),

\[(10) \quad E(\hat{B}_3) = B_1. \]

**Proof:** The proof is clear by observing

\[(11) \quad k(k-1) B_3^{(1)} = k^2 B_3^{(2)} = B_4^{(2)}. \]

Note that \( \hat{B}_3 \) may or may not be equal to \( \hat{B}_1 \). It can however be checked that, for \( k = 2, \) \( \hat{B}_3 = \hat{B}_1 = \frac{1}{2} \hat{B}_2 \).

The estimators of \( V_3 \) and \( V_4 \) are

\[(12) \quad \hat{V}_3 = \frac{k-1}{k} \sum_{u=1}^{k} (\hat{R}_1(u) - \hat{R}_1(\cdot))^2, \]

\[ \hat{V}_4 = \frac{m-1}{mk^2} \sum_{j=1}^{k} \sum_{v=1}^{m} (r_{j}(v) - R_{j}(\cdot))^2. \]

The mean square error in estimating \( R \) by \( \hat{R}_1 \) is denoted by \( \text{MSE}_1 \). We know

\[(13) \quad \text{MSE}_1 = V_1 + B_1^2, \quad i = 1, 2, 3, 4. \]

We consider \( \text{MSE}_i \) here to the second degree approximation. It can be seen in Murthy (1967) that

\[(14) \quad \text{MSE}_1 = \text{MSE}_2 = \frac{1}{k^2 x^2} \sum_{j=1}^{k} \{ \text{Var}(\hat{Y}_{j}) - 2R \text{ Cov}(\hat{Y}_{j}, \hat{X}_{j}) + R^2 \text{ Var}(\hat{X}_{j}) \}. \]
The estimators of $\text{MSE}_i$, denoted by $\hat{\text{MSE}}_i$, $i = 1, 2, 3, 4$, can be compared under any probability sampling design. We, however, consider the simple random sampling with replacement as our sampling design in subsequent sections.


We now consider the problem of comparison of $\hat{R}_i$ when the sampling design is simple random sampling with replacement. We have $\hat{X}_j = \sum_{i=1}^{m} x_{ij}/m$, $\hat{Y}_j = \sum_{i=1}^{m} y_{ij}/m$, $\hat{X}_j(v) = \sum_{i=1, i \neq v}^{m} x_{ij} / (m-1)$, and $\hat{Y}_j(v) = \sum_{i=1, i \neq v}^{m} y_{ij} / (m-1)$. It can be checked that

\begin{equation}
\sum_{v=1}^{m} \hat{X}_j(v) = m \hat{X}_j, \quad \sum_{v=1}^{m} \hat{Y}_j(v) = m \hat{Y}_j,
\end{equation}

\begin{align*}
m(k-1)\hat{X}_j(u) &= \sum_{j=1}^{k} \sum_{v=1}^{m} \hat{X}_j(v) , \\
\sum_{v=1}^{m} \hat{Y}_j(v) &= \sum_{j=1}^{k} \sum_{v=1}^{m} \hat{Y}_j(v) .
\end{align*}

3.1. Comparison of $E(\hat{B}_3)$ and $E(\hat{B}_4)$ to the second order approximation.

Denote

\begin{equation}
C = \sum_{j=1}^{k} \sum_{v=1}^{m} \sum_{v' = 1}^{m} \left[ R \text{ Cov}(\hat{X}_j(v), \hat{X}_j(v')) - \text{ Cov}(\hat{X}_j(v), \hat{Y}_j(v')) \right] .
\end{equation}

**Theorem 1.** Under (9)

\begin{equation}
k^2 X^2 \frac{m(m-1)}{m-1} E(\hat{B}_3) = k^2 \frac{m}{m-1} E(\hat{B}_4) + C .
\end{equation}

**Proof:** It can be seen from (5) and (15) that
\[(18) \quad m^2 B_4^{(2)} = m B_4^{(1)} + C.\]

It is now easy to check from (6), (8) and (18) that (17) is true. This completes the proof.

We now consider the intra-class model

\[(19) \quad \text{Var}(X_{ij}) = \sigma_x^2, \quad \text{Var}(Y_{ij}) = \sigma_y^2, \]
\[\text{Cov}(X_{ij}, Y_{ij}) = \rho_{xy} \sigma_x \sigma_y, \quad \text{Cov}(X_{ij}, X_{i',j}) = \rho_{xx} \sigma_x^2, \]
\[\text{Cov}(Y_{ij}, Y_{i',j}) = \rho_{yy} \sigma_y^2, \quad \text{Cov}(X_{ij}, Y_{i',j}) = \rho_{xy} \sigma_x \sigma_y.\]

Denote

\[(20) \quad C_1 = R \sigma_x^2 \rho_{xx} - \sigma_x \sigma_y \rho_{xy}, \quad C_2 = (R \sigma_x^2 - \rho_{xy} \sigma_x \sigma_y) - C_1.\]

**THEOREM 2.** Under (9) and (19),

\[(21) \quad k E(\hat{B}_3) = E(\hat{B}_4) + \frac{N^2}{X^2} C_1.\]

**Proof:** We get from (8) and (16)

\[(22) \quad (m-1)C = N^2 k m(m-2) C_2 + N^2 k m(m-1) C_1,\]
\[m X^2 E(\hat{B}_4) = N^2 C_2,\]
\[m X^2 k E(\hat{B}_3) = N^2 [C_2 + m C_1].\]

This completes the proof.
Corollary 1. If $\rho_{xx} = \rho'_{xy} = 0$ (i.e., $C_1 = 0$), then under (9) and (19), we have

$$k \ E(\hat{B}_3) = E(\hat{B}_4).$$

3.2. Comparison of $E(\hat{V}_3)$ and $E(\hat{V}_4)$ to the first order approximation.

We first state a proposition which is very useful in subsequent calculations.

Proposition. Let $t_1, \ldots, t_k$ be $k$ random variables with $E(t_i)$, $\text{Var}(t_i)$ and $\text{Cov}(t_i, t_j)$, $i \neq j$, being constants independent of $i$ and $j$ which are equal to $E(t)$, $\text{Var}(t)$, and $\text{Cov}(t, t')$. Then

$$E\left[\sum_{i=1}^{k} (t_i - \bar{t})^2\right] = (k-1)[\text{Var}(t) - \text{Cov}(t, t')] ,$$

where $\bar{t} = (t_1 + \ldots + t_k)/k$. It follows from (24) that

$$E(\hat{V}_3) = \frac{(k-1)^2}{k} [\text{Var}(\hat{R}_1(u)) - \text{Cov}(\hat{R}_1(u), \hat{R}(u'))]$$

$$E(\hat{V}_4) = \frac{(m-1)^2}{mn} \left[ \text{Var}(\hat{R}(v)) - \text{Cov}(\hat{R}(v), \hat{R}(v')) \right]$$

where $E(\hat{V}_3)$ and $E(\hat{V}_4)$ are constants independent of $(u, u')$ and $(v, v')$.

We find to the first order of approximation

$$\text{Cov}(\hat{R}_1(u), \hat{R}_1(u')) = \frac{1}{\chi^2} [\text{Cov}(\hat{R}(u), \hat{R}(u')) - R \text{Cov}(\hat{X}(u), \hat{Y}(u'))$$

$$- R \text{Cov}(\hat{Y}(u), \hat{X}(u')) + R^2 \text{Cov}(\hat{X}(u'), \hat{X}(u'))] .$$

Considering the model (19) with $\rho_{xx} = \rho'_{yy} = 0$ and denoting $\sigma^2 = \sigma_y^2 - 2R \rho_{xy} \sigma_x \sigma_y + R^2 \sigma_x^2$, we get to the first order approximation
$$\text{Var}(\hat{R}_1(u)) = \frac{N^2 \sigma^2}{(k-1) m X^2}$$

$$\text{Cov}(\hat{R}_1(u), \hat{R}_1(u')) = \frac{(k-2) N^2 \sigma^2}{(k-1)^2 m(m-1) X^2}$$

$$\text{Var}(r_j(v)) = \frac{N^2 \sigma^2}{x^2(m-1)}$$

$$\text{Cov}(r_j(v), r_j(v')) = \frac{N^2(m-2) \sigma^2}{x^2(m-1)^2}.$$

The following result is now easy to verify.

**Theorem 3.** Under (19) with $\rho_{xy} = \rho_y = 0$, the sampling design being simple random sampling with replacement, and to the first order of approximation, we have

$$\text{E}(\hat{V}_3) = \frac{\{(k-1)(m-1)-(k-2)\}}{m-1} \text{E}(\hat{V}_4),$$

$$\text{E}(\hat{V}_4) = V_1 = V_2.$$ 

Note that when $k = 2$, from (28), $\text{E}(\hat{V}_3) = \text{E}(\hat{V}_4)$. For $k > 2$ and $m \geq 2$, $\text{E}(\hat{V}_3) > \text{E}(\hat{V}_4)$. But for $k \geq 2$ and $m \geq 2$, $[\text{E}(\hat{B}_3)/(\text{E}(\hat{V}_3))^{1/2}] < [\text{E}(\hat{B}_4)/(\text{E}(\hat{V}_4))^{1/2}]$, where $\text{E}(\hat{B}_3)$ and $\text{E}(\hat{B}_4)$ are given in (8).

**3.3. Conclusion.**

For $k = 2$, both $\hat{R}_1$ and $\hat{R}_3$ are equally good over $\hat{R}_2$ and $\hat{R}_4$. For $k > 2$, $\hat{R}_1$ and $\hat{R}_3$ are better than $\hat{R}_2$ and $\hat{R}_4$ with respect to bias but $\hat{R}_3$ is not as good as $\hat{R}_1$, $\hat{R}_2$ and $\hat{R}_4$ with respect to first order approximation of variance.
4. **Empirical Study (to the Second Order Approximation).**

The population consists of the counties in 5 states in Mexico, namely Chiapas, Chihuahua, Guerrero, Puebla and Veracruz. We consider three different studies on the same population.

**I:** \( Y = \text{Total number of inhabitants} \)
\( X = \text{Total number of households} \)

**II:** \( Y = \text{Total number of literates} \)
\( X = \text{Total number of illiterates} \)

**III:** \( Y = \text{Total number of persons in primary activities} \)
\( X = \text{Total number of economically active people} \)

The data are obtained from the 1960 population general census in Mexico. The five states considered are very similar with respect to \( R = (Y/X) \) in Studies I, II and III. Note that in III the data in 2 counties are unavailable. We now draw 1000 times five subsamples of size thirty each by simple random sampling with replacement. In the following table we present the average values of \( \hat{R}_1, \hat{B}_1, \hat{V}_1, \hat{\text{MSE}}_1 \) for three studies. The entries in Tables A.2-A.4 are multiplied by \( 10^4 \).

In studies I, II and III, we find the average estimated bias in \( \hat{R}_4 \) is about 5, 4 and 3 times than that in \( \hat{R}_3 \). It is clear that, with respect to bias, \( \hat{R}_1 \) and \( \hat{R}_3 \) are better than \( \hat{R}_2 \) and \( \hat{R}_4 \). In comparison with respect to variance, \( \hat{R}_2 \) is superior to \( \hat{R}_1, \hat{R}_3 \) and \( \hat{R}_4 \). In studies I and III, \( \hat{R}_1 \) and \( \hat{R}_2 \) have smaller \( \hat{\text{MSE}} \). In study II, \( \hat{R}_3 \) has the largest \( \hat{\text{MSE}} \). In all studies \( \hat{R}_1 \) has the smallest \( |\hat{B}|/(\hat{V})^{\frac{1}{2}} \).
### Table A.1

The values of $N$ and $R$

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>672</td>
<td>672</td>
<td>670</td>
</tr>
<tr>
<td>$R$</td>
<td>5.39</td>
<td>1.13</td>
<td>0.67</td>
</tr>
</tbody>
</table>

### Table A.2

Average Estimated Ratio, Bias, Variance and MSE in Study I.

<table>
<thead>
<tr>
<th>Average</th>
<th>$\hat{R}_1$</th>
<th>$\hat{R}_2$</th>
<th>$\hat{R}_3$</th>
<th>$\hat{R}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{R}$</td>
<td>53914</td>
<td>53954</td>
<td>53905</td>
<td>53905</td>
</tr>
<tr>
<td>$(\hat{R}-R)^2$</td>
<td>68.562</td>
<td>78.967</td>
<td>68.321</td>
<td>73.849</td>
</tr>
<tr>
<td>$\hat{B}$</td>
<td>9.809</td>
<td>49.046</td>
<td>9.136</td>
<td>48.948</td>
</tr>
<tr>
<td>$\hat{V}$</td>
<td>90.726</td>
<td>82.200</td>
<td>94.744</td>
<td>111.590</td>
</tr>
<tr>
<td>MSE</td>
<td>91.081</td>
<td>91.081</td>
<td>95.194</td>
<td>113.384</td>
</tr>
<tr>
<td>$</td>
<td>\hat{B}</td>
<td>/\sqrt{\hat{V}}$</td>
<td>460</td>
<td>3377</td>
</tr>
</tbody>
</table>

### Table A.3

Average Estimated Ratio, Bias, Variance and MSE in Study II.

<table>
<thead>
<tr>
<th>Average</th>
<th>$\hat{R}_1$</th>
<th>$\hat{R}_2$</th>
<th>$\hat{R}_3$</th>
<th>$\hat{R}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{R}$</td>
<td>11276</td>
<td>11076</td>
<td>11334</td>
<td>11318</td>
</tr>
<tr>
<td>$(\hat{R}-R)^2$</td>
<td>408.07</td>
<td>356.02</td>
<td>429.67</td>
<td>436.65</td>
</tr>
<tr>
<td>$\hat{B}$</td>
<td>-50.35</td>
<td>-251.75</td>
<td>-57.30</td>
<td>-242.90</td>
</tr>
<tr>
<td>$\hat{V}$</td>
<td>401.02</td>
<td>375.26</td>
<td>432.59</td>
<td>377.50</td>
</tr>
<tr>
<td>MSE</td>
<td>402.10</td>
<td>402.10</td>
<td>434.07</td>
<td>389.82</td>
</tr>
<tr>
<td>$</td>
<td>\hat{B}</td>
<td>/\sqrt{\hat{V}}$</td>
<td>347</td>
<td>1846</td>
</tr>
</tbody>
</table>

### Table A.4

Average Estimated Ratio, Bias, Variance and MSE in Study III.

<table>
<thead>
<tr>
<th>Average</th>
<th>$\hat{R}_1$</th>
<th>$\hat{R}_2$</th>
<th>$\hat{R}_3$</th>
<th>$\hat{R}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{R}$</td>
<td>6752</td>
<td>7015</td>
<td>6668</td>
<td>6757</td>
</tr>
<tr>
<td>$(\hat{R}-R)^2$</td>
<td>41.200</td>
<td>37.867</td>
<td>47.889</td>
<td>46.944</td>
</tr>
<tr>
<td>$\hat{B}$</td>
<td>65.710</td>
<td>328.550</td>
<td>84.867</td>
<td>258.730</td>
</tr>
<tr>
<td>$\hat{V}$</td>
<td>40.464</td>
<td>22.220</td>
<td>52.783</td>
<td>33.782</td>
</tr>
<tr>
<td>MSE</td>
<td>41.224</td>
<td>41.224</td>
<td>54.350</td>
<td>44.330</td>
</tr>
<tr>
<td>$</td>
<td>\hat{B}</td>
<td>/\sqrt{\hat{V}}$</td>
<td>977</td>
<td>5281</td>
</tr>
</tbody>
</table>
5. Miscellaneous

Let \( r_1, \ldots, r_k \) be \( k \) estimators of \( R \) on the basis of \( k \) interpenetrating subsamples of size \( m \) each and under the same probability sampling design. Then \( \tilde{r} = \frac{r_1 + \ldots + r_k}{k} \) is an IPS estimator of \( R \). Denote

\[
\tilde{r}(u) = \frac{r_1 + \ldots + r_{u-1} + r_{u+1} + \ldots + r_k}{(k-1)}, \quad u = 1, \ldots, k
\]

and

\[
\tilde{r}(\ast) = \frac{\tilde{r}(1) + \ldots + \tilde{r}(k)}{k}.
\]

Then \( \hat{R}_5 = k\tilde{r} - (k-1)\tilde{r}(\ast) \) is another JIPS estimator of \( R \), different from \( \hat{R}_3 \) and \( \hat{R}_4 \). Clearly, \( \tilde{r}(\ast) = \tilde{r} = \hat{R}_5 \), or, in other words, the IPS estimator is identical with the JIPS estimator \( \hat{R}_5 \). It is also clear that

\[
\text{Var}(\hat{R}_5) = \frac{k-1}{k} \sum_{u=1}^{k} \left( \tilde{r}(u) - \tilde{r}(\ast) \right)^2 = \frac{1}{k(k-1)} \sum_{u=1}^{k} (r_u - \tilde{r})^2 = \text{Var}(\tilde{r}).
\]

Denote

\[
\text{Cov} = \frac{\sum_{u=1}^{k} \sum_{u' = 1}^{k} \text{Cov}(r_u, r_{u'})}{k(k-1)}
\]

(30)

which is the average of \( k(k-1) \) covariances,

\( \text{Cov}(r_u, r_{u'}), \ u \neq u', \ u, u' \) are in \{1, ..., k\}.

Under the assumption \( E(r_1) = \ldots = E(r_k) \), it follows that

\[
E[\text{Var}(\tilde{r})] = \text{Var}(\tilde{r}) - \text{Cov}.
\]

If \( \text{Cov} = 0 \), then \( \hat{\text{Var}}(\tilde{r}) \) is an unbiased estimator of \( \text{Var}(\tilde{r}) \). If \( \text{Cov} > 0 \) (\(< 0 \)), then \( \hat{\text{Var}}(\tilde{r}) \) underestimates (overestimates) \( \text{Var}(\tilde{r}) \).
Acknowledgements

The authors would like to thank Professors Jim Press (UC, Riverside), Rupert Miller and Lincoln Moses (Stanford University) for their various help during the preparation and writing of the manuscript, and to Karola Decleve for typing the paper with great care.

References


