Variance Estimation in Dual Registration
Under Population Heterogeneity

JUHA M. ALHO

ABSTRACT

The usual dual system estimator for population size can be severely biased, if there is population heterogeneity in the capture probabilities. In this note we investigate the bias of the corresponding variance estimator under heterogeneity. We show that the usual estimator is conservative, i.e., it gives too large values, if the two registration systems are negatively correlated, uncorrelated, or when the correlation is positive, but small. In the case of high positive correlation the usual estimator may yield too low values. Two alternative estimators are proposed. One is conservative under arbitrary heterogeneity. The other is conservative under Gaussian heterogeneity. The methods are applied to occupational disease data from Finland.

KEY WORDS: Capture-recapture; Dual system; Heterogeneity; Occupational diseases.

1. INTRODUCTION

Suppose there are $N$ individuals in a closed population. The problem is to estimate the unknown $N$ using dual registration. We sample twice with $n_j$ individuals captured at the $j$th time, $j = 1, 2$. Let $m_i$ be the number captured twice. Define indicator variables $u_{ji}$ and $m_i$ for $i = 1, \ldots, N$ such that $u_{ji} = 1$, if and only if individual $i$ is captured at the $j$th time only, $j = 1, 2$; and $m_i = 1$, if and only if individual $i$ is captured twice. Otherwise $u_{ji}$ and $m_i$ are zero. Define $n_{ji} = u_{ji} + m_i$ as the indicator of capture at the $j$th time, $j = 1, 2$. Let $M_i = u_{1i} + u_{2i} + m_i$ indicate capture at least once. Define the individual capture probabilities as $p_{ji} = E[n_{ji}], j = 1, 2; \text{ and } p_{12i} = E[m_i]$. Assume that the probabilities are strictly between zero and one. The fact that the probabilities are allowed to vary by individual indicates that we may have population heterogeneity in the capture probabilities. We complete the definition of the dual registration (or capture-recapture) model by assuming that the captures are independent for each individual, or $p_{12i} = p_{1i}p_{2i}$, and that the multinomial vectors

$$(u_{1i}, u_{2i}, m_i, 1 - M_i) \sim \text{Mult}(1; p_{1i}q_{2i}, p_{2i}q_{1i}, p_{1i}p_{2i}, 1 - \phi_i),$$

where $q_{ji} = 1 - p_{ji}$, $j = 1, 2$, and $\phi_i = p_{1i} + p_{2i} - p_{1i}p_{2i}$ are independent for $i = 1, \ldots, N$.

It is well-known that when capture probabilities do not vary by individual, or $p_{ji} = p_j$, $j = 1, 2$, the maximum likelihood estimator of $N$ is $\hat{N} = n_1n_2/m$ (or more precisely, the largest integer short of this value; cf., Feller 1968, p. 46). This classical estimator can be severely biased under population heterogeneity (Seber 1982, p. 565; Burnham and Overton 1979, Table 4, pp. 931-932). As shown, e.g., in Example 1 below, under homogeneous capture probabilities the asymptotic variance of $\hat{N}$ is $\text{Var}(\hat{N}) = Nq_1q_2/(p_1p_2)$, where $q_j = 1 - p_j$, $j = 1, 2$. Then $\text{Var}(\hat{N})$ can be estimated by $V_1 = n_1n_2u_1u_2/m^3$ (Sekar and Deming 1949, pp. 114-115).

---

1 Juha M. Alho, Institute for Environmental Studies and Department of Statistics, University of Illinois at Urbana-Champaign, 1101 W. Peabody Dr., Urbana IL, 61801, U.S.A.
The purpose of this note is to investigate the adequacy of the variance estimator \( V_1 \), and compare the bias of \( V_1 \) to the bias of \( \hat{N} \). One motive for investigating \( V_1 \) is that it has not been previously known whether \( V_1 \) is adequate in the case in which there is population heterogeneity, but \( \hat{N} \) is, nevertheless, consistent. This turns out to be the case. Similarly, it has not been clear when \( V_1 \) gives overestimates and thus can lead to valid confidence intervals, despite the bias of \( \hat{N} \). This turns out to be possible for one-sided intervals in special circumstances.

In Section 2 we calculate the asymptotic variance of \( \hat{N} \), as \( N \to \infty \), and derive a conservative estimator \( V_2 \) for this variance under arbitrary heterogeneity. In other words, \( V_2 \) overestimates the true asymptotic variance. One might hope that an overestimate of variance could compensate for the typically negative bias of \( \hat{N} \) and still yield valid confidence intervals. Unfortunately, this appears possible only when the bias of \( \hat{N} \) is small, or when \( N \) is small. In Section 3 the adequacy of \( V_1 \) is studied under Gaussian heterogeneity and an estimator \( V_3 \) is derived, which is conservative under this restricted type of heterogeneity. Gaussianity per se is not required for the arguments, only that the moments of the pairs \( (p_{1i}, p_{2i}) \) agree with those of a bivariate Gaussian distribution. This setup permits the ready examination of the effect of correlation between \( p_{1i} \)'s and \( p_{2i} \)'s on variance estimation, because correlation is expressible in terms of just one parameter, the ordinary moment correlation coefficient. In Section 4 we compare the bias in variance estimates to the bias of \( \hat{N} \) using empirical data relating to the registration of occupational diseases in Finland.

2. BIAS AND VARIANCE UNDER HETEROGENEITY

Define \( \rho_{jN} \) as the average probability of capture at the \( j \)th time, \( j = 1, 2 \); and let \( \rho_{12N} \) be the average of the products \( p_{1i}p_{2i} \), \( i = 1, \ldots, N \). Then, \( C_N = \rho_{12N} - \rho_{1N}\rho_{2N} \) is the covariance of the pairs \( (p_{1i}, p_{2i}) \). Assume that the limits \( \rho_{jN} \to \rho_j, j = 1, 2; \rho_{12N} \to \rho_{12}, \) and \( C_N \to C \) exist. Then we have that \( \hat{N}/N \to \rho_1 \rho_2/\rho_{12}, \) so \( \hat{N}/N - 1 \to -C/\rho_{12}, \) as \( N \to \infty \). This is the asymptotic bias of the classical estimator under population heterogeneity. Interestingly, it only depends on the first two moments of the distribution of the pairs \( (p_{1i}, p_{2i}) \). As is well-known (Sekar and Deming 1949, pp. 105-106; Seber 1982, p. 86), when the covariance is zero \( (C = 0) \), then the classical estimator is consistent; if \( C > 0 \), \( \hat{N} \) gives an underestimate; and if \( C < 0 \), it gives an overestimate. As noted above the adequacy of \( V_1 \), when the \( \rho_{j} \)'s vary from one individual to the next but still \( C = 0 \), is of particular interest.

We shall now calculate the asymptotic variance of the classical estimator under our general heterogeneity model. Note that the finite variance does not exist, because there is a positive probability that \( m = 0 \). Therefore, "asymptotic variance" properly refers here to the variance of the limiting distribution rather than to limit of the variances, as \( N \to \infty \).

**Lemma 1.** The asymptotic variance of \( \hat{N} \) is

\[
\text{Var}(\hat{N}) = \frac{N}{\rho_{12}^3} \left\{ \frac{\rho_1^4 \rho_2^2}{\rho_{12}^3} - \frac{\rho_1^4 \rho_2}{\rho_{12}^3} - \frac{\rho_1^4}{\rho_{12}^2} S_1 - \frac{\rho_1^3}{\rho_{12}^2} S_2 - \frac{\rho_1^3}{\rho_{12}} S_4 + \frac{\rho_1^2 \rho_2}{\rho_{12}^3} S_3 + \frac{\rho_1^2 \rho_2}{\rho_{12}^3} S_5 \right\}.
\]

where \( S_j = S_j/N \) for \( j = 1, \ldots, 5 \), with
\[ S_1 = \sum_{i=1}^{N} p_{1i}^2, \quad S_2 = \sum_{i=1}^{N} p_{2i}^2, \quad S_3 = \sum_{i=1}^{N} p_{1i}^2 p_{2i}, \quad S_4 = \sum_{i=1}^{N} p_{1i} p_{2i}. \]

The proof is sketched in the Appendix. We note that unlike the bias of \( \hat{N} \) that depends on the first two moments of the pairs \((p_{1i}, p_{2i})\) only, \( \text{Var}(\hat{N}) \) depends on moments up to fourth order. In special cases, such as the ones considered in Example 2 and Proposition 2, a simpler representation is possible.

**Example 1.** Suppose there is no heterogeneity in the probabilities, or \( p_{ji} = p_j, j = 1, 2 \). Then \( \bar{p}_1 = p_1, \bar{p}_2 = p_2 \); \( \bar{S}_j = p_j^2, j = 1, 2; S_3 = p_1^2 p_2^2, S_4 = p_1^2 p_2^2, \) and \( S_5 = p_1 p_2^3 \).

Hence, the asymptotic variance is \( \text{Var}(\hat{N}) = N(1 - \bar{p}_1 - \bar{p}_2 + \bar{p}_1 \bar{p}_2)/(p_1 p_2) = Nq_1 q_2/(p_1 p_2) \). Consistent estimators for \( Np_1 p_2 \) and \( Np_j \) are \( m \) and \( n_j, j = 1, 2 \). In other words, \( Np_j/n_j \rightarrow 1, j = 1, 2, \) and \( Np_1 p_2/m \rightarrow 1 \), as \( N \rightarrow \infty \). This gives us \( V_1 \) as an estimator for \( \text{Var}(\hat{N}) \).

**Example 2.** Suppose that the pairs \((p_{1i}, p_{2i}), i = 1, \ldots, N\), are independent in the sense that the distribution of \( p_{1i}'s \) is the same for each distinct value of the \( p_{2i}'s \). Then \( p_{12} = \bar{p}_1 \bar{p}_2, S_3 = \bar{S}_1 \bar{S}_2, S_4 = \bar{p}_2 S_1, S_5 = \bar{p}_1 S_2 \). Substituting into the Lemma we get

\[
\text{Var}(\hat{N}) = N \left( \frac{1}{\bar{p}_1 \bar{p}_2} - \frac{1}{\bar{p}_2} - \frac{1}{\bar{p}_1} - \frac{\bar{S}_1 \bar{S}_2}{\bar{p}_1^2 \bar{p}_2^2} + \frac{\bar{S}_1}{\bar{p}_1^2} + \frac{\bar{S}_2}{\bar{p}_2^2} \right)
\]

\[
= N \left( \frac{\bar{q}_1 \bar{q}_2}{\bar{p}_1 \bar{p}_2} - \frac{\text{cv}(p_{1i})^2}{\bar{p}_1 \bar{p}_2} \right),
\]

where \( \text{cv}(p_{ji}) = (\bar{S}_j - \bar{p}_j^2)/\bar{p}_j \) is the coefficient of variation of the \( p_{ji}'s, j = 1, 2 \). Obviously, \( \text{Var}(\hat{N}) \leq Nq_1 q_2/(p_1 p_2) \). A comparison with Example 1 shows that \( V_1 \) is a conservative estimator of \( \text{Var}(\hat{N}) \) (i.e., \( V_1 \) is asymptotically too large), when \( p_{1i}'s \) are independent of \( p_{2i}'s \). Another way of saying this is that, given the means \( \bar{p}_j, j = 1, 2 \), the largest value of the variance is obtained at homogeneity. This is analogous to the variance of the number of successes in Bernoulli trials with variable probabilities of success, cf. Feller 1968, pp. 230-231.

A comparison with Example 1 shows that \( V_1 \) is a conservative estimator of \( \text{Var}(\hat{N}) \) (i.e., \( V_1 \) is asymptotically too large), when the pairs \((p_{1i}, p_{2i})\) are independent. Note that the independence condition implies that \( C = 0 \).

When the probabilities are not independent, the classical variance estimator is not guaranteed to be conservative. A conservative estimator exists, however. It is obtained by majorizing \( \text{Var}(\hat{N}) \) by a quantity that can be estimated in terms of the observable variables. We prove in the Appendix the following general proposition.

**Proposition 1.** A conservative estimator of \( \text{Var}(\hat{N}) \) is

\[
V_2 = (n_1^2 n_2^2 + n_3^2 mu_1 + n_3^2 mu_2)/m^3,
\]

where \( u_j = n_j - m, j = 1, 2 \).
3. GAUSSIAN HETEROGEOUSITY

We shall now turn to a special case in which the sample moments of the pairs \((\rho_1, \rho_2)\), \(i = 1, \ldots, N\), agree with those of a bivariate normal, or Gaussian, distribution. This will permit a much sharper specification of a conservative variance estimator than the one obtained in the general case above. Assume that

\[
\begin{bmatrix}
X_1 \\
X_2
\end{bmatrix} \sim N
\begin{pmatrix}
\begin{bmatrix}
\mu_1 \\
\mu_2
\end{bmatrix},
\begin{bmatrix}
\nu_1^2 \mu_1^2 & \rho \nu_1 \nu_2 \mu_1 \mu_2 \\
\rho \nu_1 \nu_2 \mu_1 \mu_2 & \nu_2^2 \mu_1^2
\end{bmatrix}
\end{pmatrix},
\]

where \(|\rho| < 1\), and \(0 < \mu_j < 1\), \(j = 1, 2\). Note that \(\nu_j\)'s can be interpreted as the coefficients of variation of the distributions of \(p_{ij}\)'s. Write \(S_j = S_j/N\) for \(j = 1, \ldots, 5\), as before. Then substitute the moments of the bivariate normal distribution into Lemma 1 as follows,

\[
\bar{\rho}_j = E[X_j] = \mu_j, j = 1, 2;
\]

\[
\bar{S}_j = E[X_j^2] = \mu_j^2(1 + \nu_j^2), j = 1, 2;
\]

\[
\bar{S}_{12} = E[X_1X_2] = \mu_1\mu_2(1 + \rho \nu_1 \nu_2);
\]

\[
S_3 = E[X_1^2X_2^2] = \mu_1^2 \mu_2^2(1 + \nu_1^2 + \nu_2^2 + 4\rho \nu_1 \nu_2 + 2\rho^2 + 1)\nu_1^2 \nu_2^2;
\]

\[
S_4 = E[X_1^2X_2] = \mu_1^2 \mu_2(1 + 2\rho \nu_1 \nu_2 + \nu_1^2);
\]

\[
S_5 = E[X_1X_2^2] = \mu_1 \mu_2^2(1 + 2\rho \nu_1 \nu_2 + \nu_2^2).
\]

Straightforward, but slightly tedious calculations prove then the following proposition (details omitted).

**Proposition 2.** With the above assumptions

\[
\text{Var}(\bar{N}) = A_1A_2 + RN,
\]

where

\[
A_1 = N/(1 + \rho \nu_1 \nu_2)^2;
\]

\[
A_2 = [1 - (\mu_1 + \mu_2)(1 + \rho \nu_1 \nu_2 + \mu_1 \mu_2(1 + \rho \nu_1 \nu_2)^2)/[\mu_1 \mu_2(1 + \rho \nu_1 \nu_2)^2];
\]

\[
R = (2\rho \nu_1 \nu_2 + 3\rho^2 \nu_1^2 \nu_2^2 - \rho^2 \nu_1^2 \nu_2^2 - \rho^2 \nu_1^2 \nu_2^2 - \nu_1^2 \nu_2^2)/(1 + \rho \nu_1 \nu_2)^4.
\]

We can evaluate the classical variance estimator \(V_i = n_1n_2\mu_i \mu_2/m^3\) using this result. Note first that \(\{n_1n_2/m\}/A_1\rightarrow 1\), as \(N\rightarrow \infty\). Similarly, \(\{u_1u_2/m^2\}/A_2\rightarrow 1\). This proves the following corollary to Proposition 2: \((V_i - \text{Var}(\bar{N}))/N \rightarrow -R\), as \(N\rightarrow \infty\). For example, if \(\rho = 0\), then \(-R = \nu_1^2 \nu_2^2\), so that \(V_i\) is seen to overestimate the asymptotic variance. This is in accordance with Example 2.
How reasonable is the assumption of Gaussian moments? Certainly the capture probabilities cannot have strictly Gaussian distributions, because the Gaussian distribution always puts some probability mass outside the unit interval. On the other hand, suppose we generate the $p_{ji}$'s by taking $\logit(p_{ji}) = a_j + b_jY_{ji}$, where the pairs $(Y_{1i}, Y_{2i})$ are a sample from a bivariate normal distribution with mean zero, unit variances, and correlation $\rho$. If we have the relations $a_j = \logit(\mu_j), j = 1, 2$, and $b_j = \nu_j(1 + \mu_j)^2$, then the assumption of Gaussian moments is approximately true. In fact, even the distribution of the pairs $(p_{1i}, p_{2i})$ is in that case approximately bivariate Gaussian.

Let us consider the adequacy of $V_1$ further, under the Gaussian moments. The fact that probabilities are constrained to be between 0 and 1 means that $\mu_j$'s are between zero and one. Moreover, to be sure that most of the probability mass is in the unit square, let us assume that $0 < \nu_j \leq \frac{1}{2}, j = 1, 2$. If $\mu_j$'s are close to one, a much smaller upper bound would be needed. Assume now that $\rho \leq 0$. Then, one can show that

$$-R \geq (\rho^2\nu_1^4\nu_2^2 + \rho^2\nu_1^2\nu_2^4)/(1 + \rho\nu_1\nu_2)^4 > 0,$$

so that $V_1$ overestimates Var($\hat{N}$) for $\rho \leq 0$ also. Note that by continuity $V_1$ must overestimate Var($\hat{N}$) for some positive values of $\rho$, as well.

One can show that $R = R(\rho)$ is an increasing function of $\rho$ for at least $\rho > 0$. In the limit we have

$$-R(\rho) = (-2\nu_1\nu_2 - 2\nu_1^2\nu_2^2 + \nu_1^4\nu_2^2 + \nu_1^2\nu_2^4)/(1 + \nu_1\nu_2)^4,$$

as $\rho \to 1$. When $0 < \nu_j \leq \frac{1}{2}, j = 1, 2$, the smallest value of the above limit occurs at $\nu_1 = \nu_2 = \frac{1}{2}$. The minimum value is $-152/625 > -1/4$. Consequently, for $\rho > 0$, $V_1$ can either underestimate or overestimate Var($\hat{N}$).

The practical implications of the above results are as follows. First, if $\rho \leq 0$, then $\hat{N}$ is either consistent or it overestimates $N$ and $V_1$ gives an overestimate of the variance, so we can calculate a conservative upper confidence limit for $N$. When $\rho > 0$, $\hat{N}$ gives an underestimate of $N$. If, in addition, $\rho$ is small, then $V_1$ gives an underestimate, and we can get a conservative lower confidence limit for $N$. Obviously, these are rather special circumstances that one would not expect to be of wide practical utility.

Under the present model the asymptotic bias of $V_1$ is $> -N/4$ for all values of $\rho$. We can derive a conservative variance estimator by noting that in the Gaussian case the asymptotic relative bias of $\hat{N}$ is $-\rho\nu_1\nu_2/(1 + \rho\nu_1\nu_2) \leq -1/5$. Hence, asymptotically $5\hat{N}/4 \geq N$. A conservative estimator of Var($\hat{N}$) is, for example, $V_3 = V_1 + 5\hat{N}/16$. This can be much smaller than $V_2$ indicating that the Gaussian assumption is a very powerful one.

4. AN APPLICATION TO OCCUPATIONAL DISEASE REGISTRATION DATA

To get an idea of how large the biases may be in practice, let us look at occupational disease data from Finland as an example. The Finnish Register of Occupational Diseases has been in operation since 1964. It is kept by the Institute of Occupational Health in Helsinki. Since 1975 the number of new cases reported to the Register has varied from about 4,000 to over 7,000 annually (0.2 – 0.4 % of the employed population). Noise-induced hearing loss, diseases caused by repetitive of monotonous work (epicondylitis, bursitis, tendinivaginitis), and skin diseases
are the major diagnostic groups (cf. Vaaranen et al. 1985). The Register can be viewed as a dual registration system, because each case of disease should, under existing regulations, be reported to the Register both from the appropriate insurance company and the examining physician.

It is likely that the probability of reporting a case depends on diagnosis, for example. Indeed, based on data from the year 1981 we get the following statistics. Reports from the insurance companies, \( n_1 = 3,769 \); reports from the physicians, \( n_2 = 3,053 \); and cases reported from both sources, \( m = 1,591 \). Thus the usual dual registration estimate is \( \hat{N} = 7,232 \) with \( V_1^{0.5} = 97, V_2^{0.5} = 222, \) and \( V_3^{0.5} = 108.0 \). The closeness of \( V_3 \) to \( V_1 \) is striking. Stratifying the data into four categories by diagnosis (the three diagnostic groups mentioned above, and the remaining "other" category) yields the following estimates. Noise-induced hearing loss: \( \hat{N} = 2,230, V_1^{0.5} = 33.4, V_2^{0.5} = 47.2, \) and \( V_3^{0.5} = 42.6 \); diseases caused by repetitive or monotonous work: \( \hat{N} = 3,572, V_1^{0.5} = 201.4, V_2^{0.5} = 303.8, \) and \( V_3^{0.5} = 204.2 \); skin diseases: \( \hat{N} = 1,441, V_1^{0.5} = 30.9, V_2^{0.5} = 86.2, \) and \( V_3^{0.5} = 37.5 \); other diseases \( \hat{N} = 1,015, V_1^{0.5} = 32.7, V_2^{0.5} = 79.1, \) and \( V_3^{0.5} = 37.2 \). Adding the results yields the following estimates for the total number of diseases: \( \hat{N} = 8,258, V_1^{0.5} = 209.0, V_2^{0.5} = 340.3, \) and \( V_3^{0.5} = 215.2 \). We see that diseases caused by repetitive or monotonous work are underreported to a particularly great extent.

The analysis was extended further by stratifying the data by diagnosis (4 categories), insurance company (11 categories), and main groups of industry (7 categories). \( A \) priori, these factors could be thought to have an influence on reporting probabilities. However, the stratification did not alter the point estimate materially. It did increase the estimated standard deviations by over a third, apparently because some of the strata became very small. We conclude that the bias in the point estimator caused by diagnosis is the dominant source of error in the classical estimator in this application.

The same data were further analyzed using a logistic regression technique that allows us to take into account observable population heterogeneity due to both discrete and continuous explanatory variables. In this application age was shown to have an effect on reporting probabilities within the diagnostic groups for one source of information, but not for the other. Therefore, the point estimates remained unchanged and the conclusion regarding the role of diagnosis could not be refuted (Alho 1990).

5. DISCUSSION

Our theoretical results indicate that the usual variance estimator \( V_1 \) is conservative when the two registration systems are negatively correlated or independent. By continuity the estimator may be conservative also when the correlation is positive but small. Under high positive correlation \( V_1 \) gives too low values. We introduced an alternative estimator \( V_2 \), which is conservative under arbitrary population heterogeneity. However, it appears to be unduly conservative in view of the numerical comparisons with \( V_3 \), which is guaranteed to be conservative under Gaussian heterogeneity. The closeness of \( V_3 \) to \( V_1 \) suggests that, in practice, \( V_1 \) may be fairly robust against population heterogeneity.

Unfortunately, even the use of the conservative estimator \( V_2 \) would not have been sufficient to cover the bias in the classical point estimator in our empirical example. Perhaps this was to be expected, since the bias of \( \hat{N} \) and the degree of overestimation provided by \( V_2 \) are both of order \( N \). Hence, the use of \( V_2 \) inflates the width of a confidence interval by a factor of order \( N^{0.5} \) only. Therefore, \( V_2 \) can compensate for the bias of \( \hat{N} \), if the bias is small, or if \( N \) itself is small. Hence, it seems that the successful application of the dual registration method requires that either we have roughly uncorrelated registration systems, or that the heterogeneity is
observable. In the latter case we may use stratification as suggested already by Sekar and Deming (1949), or logistic regression modeling as suggested by Huggins (1989) and Alho (1990), to adjust for the bias of the classical estimator of population size.

ACKNOWLEDGEMENT

The author would like to thank Bruce Spencer and an anonymous referee for comments that helped to improve the presentation. Part of the empirical results were first presented at the 11th Nordic Conference on Mathematical Statistics in Uppsalá, Sweden, in June 1986.

APPENDIX

Proof of Lemma 1. Apply a linear Taylor-series development to \( \hat{N} = n_1n_2/m \) at \( \bar{N} \)
\[
E \left[ \frac{n_1 n_2}{m} \right] = N\bar{p}_1\bar{p}_2/\bar{p}_{12},
\]
or
\[
\hat{N} = \frac{N\bar{p}_1\bar{p}_2}{\bar{p}_{12}} + \frac{\bar{p}_2}{\bar{p}_{12}} (n_1 - N\bar{p}_1) + \frac{\bar{p}_1}{\bar{p}_{12}} (n_2 - N\bar{p}_2) - \frac{\bar{p}_1\bar{p}_2}{\bar{p}_{12}^2} (m - N\bar{p}_{12}).
\]

Hence, we have
\[
E \left[ \left( \frac{\hat{N} - N\bar{p}_1\bar{p}_2}{\bar{p}_{12}} \right)^2 \right] = \left( \frac{\bar{p}_2}{\bar{p}_{12}} \right)^2 \text{Var}(n_1) + \left( \frac{\bar{p}_1}{\bar{p}_{12}} \right)^2 \text{Var}(n_2) + \left( \frac{\bar{p}_1\bar{p}_2}{\bar{p}_{12}^2} \right)^2 \text{Var}(m)
\]
\[
- \frac{\bar{p}_1\bar{p}_2}{\bar{p}_{12}} \text{Cov}(n_1, m) - 2\frac{\bar{p}_1\bar{p}_2}{\bar{p}_{12}^2} \text{Cov}(n_2, m).
\]

Under our independence assumptions \( \text{Var}(n_j) = N\bar{p}_j - S_j, j = 1, 2; \text{Var}(m) = N\bar{p}_{12} - S_3, \text{Cov}(n_1, m) = -S_4 + N\bar{p}_{12}, \) and \( \text{Cov}(n_2, m) = -S_5 + N\bar{p}_{12}. \) Substituting these into the mean squared error gives the result.

Proof of Proposition 1. We ignore the negative term containing \( S_3 \) in Lemma 1. Since \( 0 < p_{ji} < 1 \), we have \( S_4 < N\bar{p}_{12} \), and \( S_4 < S_1 \). Therefore,
\[
\frac{2\bar{p}_1\bar{p}_2}{\bar{p}_{12}^3} S_4 < \frac{\bar{p}_1\bar{p}_2}{\bar{p}_{12}} N\bar{p}_{12} + \frac{\bar{p}_2}{\bar{p}_{12}} S_1 + \left( \frac{\bar{p}_1 - \bar{p}_{12}}{\bar{p}_{12}} \right) \frac{\bar{p}_2^2}{\bar{p}_{12}^2} N\bar{p}_{12}.
\]

Similarly,
\[
\frac{2\bar{p}_1\bar{p}_2}{\bar{p}_{12}^3} S_5 < \frac{\bar{p}_1\bar{p}_2}{\bar{p}_{12}} N\bar{p}_{12} + \frac{\bar{p}_1}{\bar{p}_{12}} S_2 + \left( \frac{\bar{p}_2 - \bar{p}_{12}}{\bar{p}_{12}} \right) \frac{\bar{p}_1^2}{\bar{p}_{12}^2} N\bar{p}_{12}.
\]

Substituting these bounds to the expression of Lemma 1 we get
\[
\text{Var}(\hat{N}) < \frac{\bar{p}_2^2}{\bar{p}_{12}^3} N + \left( \frac{\bar{p}_1 - \bar{p}_{12}}{\bar{p}_{12}} \right) \frac{\bar{p}_2^2}{\bar{p}_{12}^2} N + \left( \frac{\bar{p}_2 - \bar{p}_{12}}{\bar{p}_{12}} \right) \frac{\bar{p}_1^2}{\bar{p}_{12}^2} N.
\]

Estimating \( N\bar{p}_j \) by \( n_j, j = 1, 2; \) and \( N\bar{p}_{12} \) by \( m \) we get the result.
REFERENCES


