Maximum-Likelihood Estimate of Proportion using Supplementary Information

Saito, Kin-ichiro
Sophia University

http://hdl.handle.net/2324/12973

出版情報：統計数理研究. 7 (1/2), pp.11-17, 1956-12. 統計科学研究会
バージョン: published
権利関係：
MAXIMUM-LIKELIHOOD ESTIMATE OF PROPORTION USING SUPPLEMENTARY INFORMATION

By

Kin-ichirō Saitō

(Received Oct. 20, 1956)

1. Introduction. When a random sample is taken from a population, the proportion of units in the sample possessing certain characteristic provides an estimate of the corresponding proportion of units in the population. It is well known that this estimate is the best one which can be derived from the sample data independently of any supplementary information associated with the characteristic about which information is required. By the use of such supplementary information as is available, however, further alternative estimates of the population proportion can usually be derived that are more accurate than the simple estimate without use of this information.

Supplementary information takes often in practice the form of a known proportion of units in the population possessing some characteristic auxiliary to the characteristic under investigation. The purpose of this paper is to derive the maximum-likelihood estimate of the proportion in question under this situation, and to compare its precision in the case of large sample with that of the simple estimate which does not utilize the supplementary information.

For convenience of discussions, the following four cases will be distinguished with regards the relation between the characteristic under investigation and the auxiliary characteristic possessed by a known proportion of units in the population.

(a) The case in which the characteristic under investigation is conceptually subordinate to the auxiliary characteristic. In this case, the set of all units with the auxiliary characteristic includes the set of all units with the characteristic under investigation.

(b) The case in which the auxiliary characteristic is conceptually subordinate to the characteristic under investigation. In this case, the set of all units with the characteristic under investigation includes the set of all units with the auxiliary characteristic.

(c) The case in which the characteristic under investigation and the auxiliary characteristic are conceptually exclusive to each other. In this case, the set of all units with the characteristic under investigation has no common part with the set of all units possessing the auxiliary characteristic.
(d) The general case in which there is no relation of subordination or mutual exclusiveness between the characteristic under investigation and the auxiliary characteristic. In this case, the set of all units with the characteristic under investigation has a common part with the set of all units possessing the auxiliary characteristic.

Our problem of estimation will be considered for these cases in turn. The following notation will be used in the sequel.

- \( X \), set of units in the population possessing the characteristic under investigation.
- \( X^* \), complementary set of \( X \), i.e., set of units in the population which do not possess the characteristic under investigation.
- \( A \), set of units in the population possessing the auxiliary characteristic.
- \( A^* \), complementary set of \( A \), i.e., set of units in the population which do not possess the auxiliary characteristic.
- \( P_x \), proportion of unit in the population possessing the characteristic under investigation.
- \( Q_x = 1 - P_x \), proportion of units in the population possessing the auxiliary characteristic.
- \( Q_A = 1 - P_A \), proportion of units in the population possessing the auxiliary characteristic.
- \( n \), size of the sample.
- \( x \), number of units in the sample possessing the characteristic under investigation.
- \( a \), number of units in the sample possessing the auxiliary characteristic.
- \( \hat{p}_x = x/n \), \( q_x = 1 - \hat{p}_x \).
- \( \hat{p}_A = a/n \), \( q_A = 1 - \hat{p}_A \).

2. The case in which the characteristic under investigation is conceptually subordinate to the auxiliary characteristic. In this case, \( A \) includes \( X \) as its subset. We shall assume that the sample size \( n \) is sufficiently small compared with the population size for us to be able to ignore the complications of sampling without replacement. The likelihood is, then

\[ e^i = P_x^a (P_A - P_x)^{a-x} (1 - P_A)^{n-a}. \]

Therefore

\[ L = x \log P_x + (a - x) \log (P_A - P_x) + (n - a) \log (1 - P_A). \]

Differentiating with respect to \( P_x \) gives

\[ \frac{\partial L}{\partial P_x} = \frac{x}{P_x} - \frac{a - x}{P_A - P_x}. \]
Hence the maximum-likelihood estimate of $P_x$ is

\[(2.3) \hat{P}_x = \frac{P_a}{a} x.\]

The conditional expectation of $\hat{P}_x$ when $a$ is kept fixed is obviously equal to $P_x$. Therefore, $\hat{P}_x$ is an unbiased estimate of $P_x$. This estimate $\hat{P}_x$ represents what is called ratio estimate for a qualitative characteristic.

Differentiating (2.2) with respect to $P_x$, and then taking the expectation provides the information in this case of the sample. Thus

\[(2.4) I_x = -E \frac{\partial^2 L}{\partial P_x^2} = \frac{n P_a}{P_x (P_a - P_x)}.\]

Hence, the asymptotic variance of $\hat{P}_x$ is

\[(2.5) \text{Var} \hat{P}_x = \frac{P_x}{n} \left( 1 - \frac{P_x}{P_a} \right).\]

Using the value of $\hat{P}_x$ given by (2.3) in place of $P_x$ in (2.5), $\text{Var} \hat{P}_x$ may be estimated by

\[(2.6) \text{Var} \hat{P}_x = \frac{P_a}{n} \cdot \frac{x}{a} \left( 1 - \frac{x}{a} \right).\]

The variance of the simple estimate $p_x$ without use of the supplementary information is

\[(2.7) \text{Var} p_x = \frac{1}{n} P_x Q_x.\]

Let us denote by $d$ the gain in precision of $\hat{P}_x$ over $p_x$, then

\[1 + d = \frac{\text{Var} p_x}{\text{Var} \hat{P}_x} = \frac{Q_x}{1 - \frac{P_x}{P_a}}.\]

Therefore

\[(2.8) d = \frac{P_x Q_a}{P_a - P_x} \geq 0.\]

If we put $\lambda = P_x/P_a$, then

\[d = \frac{\lambda}{1 - \lambda} Q_a.\]

When $\lambda$ is large and $P_a$ is small, the gain in precision is considerable. For example, if $\lambda = 0.9$, $P_a = 0.01$ then $d = 8.9$. 

3. The case in which the auxiliary characteristic is conceptually subordinate to the characteristic under investigation. In this case the set $X$ includes the set $A$ as its subset.

Since $X$ includes $A$, $A^*$ includes $X^*$. Hence, the maximum-likelihood estimate of $Q_x$ may be obtained using $Q_A$ by the method described in 2. Thus

\[(3.1) \hat{Q}_x = \frac{Q_A}{n-a} (n-x) = \frac{Q_A q_x}{q_A}.
\]

Therefore the maximum-likelihood estimate of $P_x$ is

\[(3.2) \hat{P}_x = 1 - \frac{Q_A q_x}{q_A} = p_x + q_x \left(1 - \frac{Q_A}{q_A}\right).
\]

This $\hat{P}_x$ is an unbiased estimate of $P_x$. The asymptotic variance of $\hat{P}_x$ is

\[(3.3) \text{Var} \hat{P}_x = \frac{1}{n} Q_x \left(1 - \frac{Q_A}{Q_x}\right).
\]

Using $\hat{Q}_x$ given by (3.1) in place of $Q_x$ in (3.3), $\text{Var} \hat{P}_x$ may be estimated by

\[(3.4) \nu(\hat{P}_x) = \frac{Q_A q_x}{n q_A} \left(1 - \frac{q_x}{q_A}\right).
\]

If we denote by $d$ the gain in precision of $\hat{P}_x$ over $p_x$, the simple estimate without use of the supplementary information

\[1 + d = \frac{\text{Var} \hat{p}_x}{\text{Var} \hat{P}_x} = \frac{P_x Q_A}{Q_A - Q_x}.
\]

Therefore

\[(3.5) d = \frac{P_x Q_A}{P_x - P_A} = \frac{\lambda}{1 - \lambda} Q_x \geq 0, \quad \frac{P_A}{P_x} = \lambda, \quad \text{say.}
\]

When $\lambda$ is large and $P_x$ is small, the gain in precision is considerable.

4. The case in which the characteristic under investigation and the auxiliary characteristic are conceptually exclusive to each other. From the relation of the two characteristics, we see that the sets $A$ and $B$ are exclusive. Since $X$ has no common part with $A$, $A^*$ includes $X$.

Hence the maximum-likelihood estimate of $P_x$ may be had using $Q_A$ by the method developed in 2. Thus

\[(4.1) \hat{P}_x = \frac{Q_A}{n-a} x = \frac{Q_A q_x}{q_A} p_x.
\]
This is an unbiased estimate of $P_x$. The asymptotic variance of $\hat{P}_x$ is

\begin{equation}
\text{Var} \hat{P}_x = \frac{1}{n} P_x \left(1 - \frac{P_x}{Q_A}\right).
\end{equation}

Using $\hat{P}_x$ given by (4.1) in place of $P_x$ in (4.2), an estimate of $\text{Var} \hat{P}_x$ is obtained as follows.

\begin{equation}
\nu(\hat{P}_x) = \frac{Q_A \hat{P}_x}{n} \left(1 - \frac{\hat{P}_x}{Q_A}\right).
\end{equation}

If we denote by $d$ the gain in precision of $\hat{P}_x$ over $P_x$

\begin{equation}
1 + d = \frac{\text{Var} P_x}{\text{Var} \hat{P}_x} = \frac{Q_A Q_x}{Q_A - P_x}.
\end{equation}

Therefore

\begin{equation}
d = \frac{P_x P_A}{1 - P_A - P_x} \geq 0.
\end{equation}

When $P_A$ and $P_x$ together constitute a large part of the population, the gain in precision may be substantial. For example, if $P_A = 0.5$, $P_x = 0.4$, then $d = 2.0$.

5. The general case. In this case $A$ and $X$ are not disjoint and neither of them includes the other.

Let $Y$ be the common part of $A$ and $X$, and $P_y$ be the corresponding proportion. Let $Z$ be $X - Y$, and $P_z$ be the corresponding proportion. Since $A$ includes $Y$, the maximum-likelihood estimate of $P_y$ may be had by the method described in 2. Since $A$ and $Z$ have no common part, the maximum-likelihood estimate of $P_z$ may be had by the method described in 4. Thus

\begin{equation}
\hat{P}_y = \frac{P_A}{a} y = \frac{P_A}{P_y} p_y,
\end{equation}

\begin{equation}
\hat{P}_z = \frac{Q_A}{n - a} Z_A = \frac{Q_A}{P_z} p_z,
\end{equation}

where $y$ is the number of units in the sample possessing the both characteristics, and $z$ is the number of units in the sample possessing the characteristic under investigation but not the auxiliary characteristic, $p_x$ and $p_z$ representing $y/n$ and $z/n$ respectively.

Since $P_x = P_y + P_z$, the maximum-likelihood estimate of $P_x$ is

\begin{equation}
\hat{P}_x = \hat{P}_y + \hat{P}_z = \frac{P_A}{P_y} p_y + \frac{Q_A}{P_z} p_z.
\end{equation}

This is an unbiased estimate of $P_x$. 
The variance of $\hat{P}_x$ may be expressed as follows,

$$\text{Var} \hat{P}_x = \text{Var} \hat{P}_r + \text{Var} \hat{P}_z + 2 \text{Cov}(\hat{P}_r, \hat{P}_z).$$  

The likelihood, in this case, is expressed as

$$e^L = P_r^y P_z^x (P_A - P_r)^{a-y} (1 - P_A - P_x)^{n-a-x}. $$

Therefore

$$L = y \log P_r + z \log P_z + (a - y) \log (P_A - P_r) + (n - a - z) \log (1 - P_A - P_x).$$

The information matrix is then

$$I = \begin{pmatrix} -E \frac{\partial^2 L}{\partial P_r^2} & -E \frac{\partial^2 L}{\partial P_r \partial P_z} \\ -E \frac{\partial^2 L}{\partial P_r \partial P_z} & -E \frac{\partial^2 L}{\partial P_z^2} \end{pmatrix} = \begin{pmatrix} \frac{n P_r}{P_r (P_A - P_r)} & 0 \\ 0 & \frac{n (1 - P_A)}{P_z (1 - P_A - P_z)} \end{pmatrix}.$$ 

Therefore the asymptotic variance-covariance matrix of $\hat{P}_r$ and $\hat{P}_z$ is

$$I^{-1} = \begin{pmatrix} \frac{1}{n} P_r \left(1 - \frac{P_r}{P_A}\right) & 0 \\ 0 & \frac{1}{n} P_z \left(1 - \frac{P_z}{Q_A}\right) \end{pmatrix}. $$

Hence for the case of large sample

$$\text{Var} \hat{P}_r = \frac{1}{n} P_r \left(1 - \frac{P_r}{P_A}\right),$$

$$\text{Var} \hat{P}_z = \frac{1}{n} P_z \left(1 - \frac{P_z}{Q_A}\right),$$

$$\text{Cov}(\hat{P}_r, \hat{P}_z) = 0.$$

Inserting (5.9), (5.10), (5.11) in (5.4), we obtain

$$\text{Var} \hat{P}_x = \frac{1}{n} \left[ P_r \left(1 - \frac{P_r}{P_A}\right) + P_z \left(1 - \frac{P_z}{Q_A}\right) \right].$$

Using the values of $\hat{P}_r$ and $\hat{P}_x$ given by (5.1) and (5.2) in place of $P_r$ and $P_x$ in (5.12), an estimate of $\text{Var} \hat{P}_x$ may be had as follows,

$$V(\hat{P}_x) = \frac{1}{n} \left[ P_A \frac{P_x}{\hat{P}_a} \left(1 - \frac{\hat{P}_r}{\hat{P}_a}\right) + Q_A \frac{P_z}{\hat{q}_a} \left(1 - \frac{\hat{p}_x}{\hat{q}_a}\right) \right].$$

Denoting the gain in precision of $\hat{P}_x$ over the simple estimate $\hat{p}_x$ by $d$, we have
(5.14) \[ 1 + d = \frac{\text{Var} \hat{P}_x}{\text{Var} \hat{P}_x} = \frac{(P_x + P_z)(1 - P_x - P_z)}{P_x(1 - P_x) + P_z(1 - P_z)}. \]

From this, we get

(5.15) \[ d = \frac{(P_x \sqrt{\frac{Q}{P_A}} - P_z \sqrt{\frac{P_A}{Q}})^2}{P_x(1 - \frac{P_x}{P_A}) + P_z(1 - \frac{P_z}{Q_A})}. \]

Since \( P_x \leq P_A, P_z \leq Q_A \), \( d \) is never negative.

It may be observed that

\[ d = 0 \]

when and only when

(5.16) \[ \frac{P_x}{P_z} = \frac{P_A}{Q_A}. \]

The equation (5.16) is equivalent to the relation

(5.17) \[ P_x = P_A \cdot P_x, \]

which represents the independence of the two characteristics.

The gain in precision of \( \hat{P}_x \) over \( \hat{P}_x \) can be considerable. For example, if \( P_A = 0.1, P_x = 0.09 \) and \( P_z = 0.01 \), then \( d = 3.6 \).

6. Remarks. It may be easily verified that the estimates derived above are equivalent to those obtained by what F. YATES calls the stratification after selection. The argument so far developed shows that these have the asymptotic optimum properties of maximum-likelihood estimates.

SOPHIA UNIVERSITY

Reference