Maximum-Likelihood Estimate of Proportion using Supplementary Information

Saito, Kin-ichiro
Sophia University

https://doi.org/10.5109/12973

出版情報：統計数理研究. 7 (1/2), pp.11-17, 1956-12. Research Association of Statistical Sciences

MAXIMUM-LIKELIHOOD ESTIMATE OF PROPORTION USING SUPPLEMENTARY INFORMATION

By

Kin-ichirô SAITÔ

(*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 chacteristic are conceptually exclusive to each other. In this case, the set of all units with the characteristic under investigation has no commun 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 chararcteristic.
P_X, proportion of unit in the population possessing the characteristic under investigation.
$Q_X = 1 - P_X$.
P_A, proportion of units in the population possessing the auxiliary characteristic.
$Q_A = 1 - P_A$.
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.
$p_X = x/n, \quad q_X = 1 - p_X,$
$p_A = a/n, \quad q_A = 1 - p_A.$

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

$$e^L = P_X^x (P_A - P_X)^{a-x} (1 - P_A)^{n-a}.$$

Therefore

(2.1) $$L = x \log P_X + (a - x) \log (P_A - P_X) + (n - a) \log (1 - P_A).$$

Differentiating with respect to P_X gives

(2.2) $$\frac{\partial L}{\partial P_X} = \frac{x}{P_X} - \frac{a - x}{P_A - P_X}.$$

Hence the maximum-likelihood estimate of P_x is

$$\hat{P}_x = \frac{P_A}{a} x . \tag{2.3}$$

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

$$I_x = -E\frac{\partial^2 L}{\partial P_x^2} = \frac{nP_A}{P_x(P_A - P_x)} . \tag{2.4}$$

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

$$\text{Var}\,\hat{P}_x = \frac{P_x}{n}\left(1 - \frac{P_x}{P_A}\right) . \tag{2.5}$$

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

$$v(\hat{P}_x) = \frac{P_A}{n} \cdot \frac{x}{a}\left(1 - \frac{x}{a}\right) . \tag{2.6}$$

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

$$\text{Var}\,p_x = \frac{1}{n} P_x Q_x . \tag{2.7}$$

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

$$d = \frac{P_x Q_A}{P_A - P_x} \geqq 0 . \tag{2.8}$$

If we put $\lambda = P_x/P_A$,

$$d = \frac{\lambda}{1 - \lambda} Q_A .$$

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

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

$$\hat{Q}_X = \frac{Q_A}{n-a}(n-x) = \frac{Q_A}{q_A} q_X . \tag{3.1}$$

Therefore the maximum-likelihood estimate of P_X is

$$\hat{P}_X = 1 - \frac{Q_A}{q_A} q_X = p_X + q_X\left(1 - \frac{Q_A}{q_A}\right). \tag{3.2}$$

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

$$\operatorname{Var} \hat{P}_X = \frac{1}{n} Q_X\left(1 - \frac{Q_X}{Q_A}\right). \tag{3.3}$$

Using $\hat{Q}_X$ given by (3.1) in place of Q_X in (3.3), $\operatorname{Var}\hat{P}_X$ may be estimated by

$$v(\hat{P}_X) = \frac{Q_A}{n}\frac{q_X}{q_A}\left(1 - \frac{q_X}{q_A}\right). \tag{3.4}$$

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

$$1 + d = \frac{\operatorname{Var} p_X}{\operatorname{Var} \hat{P}_X} = \frac{P_X Q_A}{Q_A - Q_X}.$$

Therefore

$$d = \frac{P_A Q_X}{P_X - P_A} = \frac{\lambda}{1-\lambda} Q_X \geqq 0, \quad \frac{P_A}{P_X} = \lambda, \quad \text{say.} \tag{3.5}$$

When λ 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 characteristecs, 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

$$\hat{P}_X = \frac{Q_A}{n-a} x = \frac{Q_A}{q_A} p_X . \tag{4.1}$$

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

$$\text{(4.2)} \qquad \operatorname{Var}\hat{P}_X=\frac{1}{n}P_X\left(1-\frac{P_X}{Q_A}\right).$$

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

$$\text{(4.3)} \qquad v(\hat{P}_X)=\frac{Q_A}{n}\frac{p_X}{q_A}\left(1-\frac{p_X}{q_A}\right).$$

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

$$1+d=\frac{\operatorname{Var}p_X}{\operatorname{Var}\hat{P}_X}=\frac{Q_AQ_X}{Q_A-P_X}.$$

Therefore

$$\text{(4.4)} \qquad d=\frac{P_AP_X}{1-P_A-P_X}\geqq 0.$$

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

$$\text{(5.1)} \qquad \hat{P}_Y=\frac{P_A}{a}y=\frac{P_A}{p_A}p_Y,$$

$$\text{(5.2)} \qquad \hat{P}_Z=\frac{Q_A}{n-a}z_A=\frac{Q_A}{q_A}p_Z,$$

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_Y 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

$$\text{(5.3)} \qquad \hat{P}_X=\hat{P}_Y+\hat{P}_Z=\frac{P_A}{p_A}p_Y+\frac{Q_A}{q_A}p_Z.$$

This is an unbiased estimate of P_X.

The variance of $\hat{P}_X$ may be expressed as follows,

$$\text{(5.4)} \qquad \operatorname{Var}\hat{P}_X = \operatorname{Var}\hat{P}_Y + \operatorname{Var}\hat{P}_Z + 2\operatorname{Cov}(\hat{P}_Y, \hat{P}_Z).$$

The likelihood, in this case, is expressed as

$$\text{(5.5)} \qquad e^L = P_Y^y P_Z^z (P_A - P_Y)^{a-y}(1 - P_A - P_Z)^{n-a-z}.$$

Therefore

$$\text{(5.6)} \qquad L = y \log P_Y + z \log P_Z + (a-y)\log(P_A - P_Y) + (n-a-z)\log(1 - P_A - P_Z).$$

The information matrix is then

$$\text{(5.7)} \qquad I = \begin{pmatrix} -E\dfrac{\partial^2 L}{\partial P_Y^2} & -E\dfrac{\partial^2 L}{\partial P_Y \partial P_Z} \\ -E\dfrac{\partial^2 L}{\partial P_Y \partial P_Z} & -E\dfrac{\partial^2 L}{\partial P_Z^2} \end{pmatrix} = \begin{pmatrix} \dfrac{nP_A}{P_Y(P_A - P_Y)} & 0 \\ 0 & \dfrac{n(1-P_A)}{P_Z(1-P_A-P_Z)} \end{pmatrix}.$$

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

$$\text{(5.8)} \qquad I^{-1} = \begin{pmatrix} \dfrac{1}{n}P_Y\left(1 - \dfrac{P_Y}{P_A}\right) & 0 \\ 0 & \dfrac{1}{n}P_Z\left(1 - \dfrac{P_Z}{Q_A}\right) \end{pmatrix}.$$

Hence for the case of large sample

$$\text{(5.9)} \qquad \operatorname{Var}\hat{P}_Y = \frac{1}{n} P_Y\left(1 - \frac{P_Y}{P_A}\right),$$

$$\text{(5.10)} \qquad \operatorname{Var}\hat{P}_Z = \frac{1}{n} P_Z\left(1 - \frac{P_Z}{Q_A}\right),$$

$$\text{(5.11)} \qquad \operatorname{Cov}(P_Y, P_Z) = 0.$$

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

$$\text{(5.12)} \qquad \operatorname{Var}\hat{P}_X = \frac{1}{n}\left[P_Y\left(1 - \frac{P_Y}{P_A}\right) + P_Z\left(1 - \frac{P_Z}{Q_A}\right)\right].$$

Using the values of $\hat{P}_Y$ and $\hat{P}_Z$ given by (5.1) and (5.2) in place of P_Y and P_Z in (5.12), an estimate of $\operatorname{Var}\hat{P}_X$ may be had as follows,

$$\text{(5.13)} \qquad V(\hat{P}_X) = \frac{1}{n}\left[P_A\frac{p_Y}{p_A}\left(1 - \frac{p_Y}{p_A}\right) + Q_A\frac{p_Z}{q_A}\left(1 - \frac{p_Z}{q_A}\right)\right].$$

Denoting the gain in precision of $\hat{P}_X$ over the simple estimate p_X by d, we have

(5.14) $$1+d=\frac{\mathrm{Var}\,\hat{p}_X}{\mathrm{Var}\,\hat{P}_X}=\frac{(P_Y+P_Z)(1-P_Y-P_Z)}{P_Y\left(1-\frac{P_Y}{P_A}\right)+P_Z\left(1-\frac{P_Z}{Q_A}\right)}.$$

From this, we get

(5.15) $$d=\frac{\left(P_Y\sqrt{\frac{Q_A}{P_A}}-P_Z\sqrt{\frac{P_A}{Q_A}}\right)^2}{P_Y\left(1-\frac{P_Y}{P_A}\right)+P_Z\left(1-\frac{P_Z}{Q_A}\right)}.$$

Since $P_Y \leqq P_A$, $P_Z \leqq Q_A$, d is never negative.
It may be observed that

$$d=0$$

when and only when

(5.16) $$\frac{P_Y}{P_Z}=\frac{P_A}{Q_A}.$$

The equation (5.16) is equivalent to the relation

(5.17) $$P_Y=P_A\cdot P_X,$$

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 $P_A=0.1$, $P_Y=0.09$ 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

F. YATES; Sampling methods for censuses and surveys. Griffin, London, 1949.