Sampling plan using grade of infestation to estimate population density

Kohji YAMAMURA¹,* and Hisashi NEMOTO²

¹ Laboratory of Population Ecology, National Institute for Agro-Environmental Sciences; Tsukuba 305-8604, Japan
² Saitama Prefectural Agriculture and Research Center; Kuki, Saitama 346-0037, Japan

(Received 6 November 2001; Accepted 13 December 2002)

Abstract

By a grading method, we classify the infestation level into grades depending on the number of individuals and record the grade of infestation without recording the number of individuals. However, the resultant grade has no clear meaning and hence it currently has only limited use. There are two questions that should be solved to construct a grading method that yields a meaningful grade. (1) Which scale should we use to determine the boundary of grades? For example, is a logarithmic scale such as that divided at $x=1, 10, 100, 1000$ preferable to an arithmetic scale such as that divided at $x=50, 100, 150, 200$? (2) Which grade width should we use? When we use a logarithmic scale, for example, is a 5-times width such as that divided at $x=1, 5, 25, 125$ preferable to a 10-times width such as that divided at $x=1, 10, 100, 1000$? For the first question, we suggest that we should utilize the scale that improves homoscedasticity to perform ANOVA. For the second question, we provide a procedure to determine the optimal grade width.

Key words: Bias; homoscedasticity; mean square error; accuracy; variance

INTRODUCTION

It is usually laborious to count the number of individuals of gregarious small organisms such as aphids and mites in the field. In such a case, a grading method is frequently used by which the number of individuals is classified into one of several grades. The grade is recorded to describe the infestation level of each sampling unit, that is, without recording the exact number of individuals. In estimating the density of citrus rust mite, *Phyllocoptruta oleivora* (Ashmead), Rogers et al. (1994) used a grading method based on a modified Horsfall-Barratt system, which was originally used for the assessment of plant diseases. They classified the infestation level into 7 grades depending on the number of individuals, $x$ in the sample, as follows: $0$, $1 \leq x \leq 3$, $4 \leq x \leq 6$, $7 \leq x \leq 12$, $13 \leq x \leq 25$, $26 \leq x \leq 50$, and $51 \leq x$. Similar methods have been used in all prefectures in Japan from the 1950s until today in application of the Japanese plant protection standards for evaluating the infestation level of various insect pests (Ministry of Agriculture Forestry and Fisheries, 1986). Five grades are usually used in these standards: none, low, medium, high, and very high. In Japanese, these grades are called '甚', '多', '中', '少', '無' in descending order. When we examine the number of green rice leafhopper, *Nephotettix cincticeps* (Uhler) (Deltocephalidae; Hemiptera), in paddy fields after transplanting by sweeping the field 20 times with a net, for example, the infestation level is recorded as the following grades: none: $x=0$, low: $1 \leq x \leq 50$, medium: $51 \leq x \leq 750$, high: $751 \leq x \leq 1,500$, and very high: $1,501 \leq x$. The original Horsfall-Barratt system for estimating the percentage of plant diseases adopted a modified logarithmic scale (Horsfall and Barratt, 1945). Below 50%, the eye sees the amount of diseased tissue in the system. Above 50%, it sees the amount of disease-free tissue. Then, the percentage of plant disease, $p$, is described by twelve grades: $p=0$, $0 < p < 3$, $3 \leq p < 6$, $6 \leq p < 12$, $12 \leq p < 25$, $25 \leq p < 50$, $50 \leq p < 75$, $75 \leq p < 87$, $87 \leq p < 94$, $94 \leq p < 97$, $97 \leq p < 100$, and $p=100$. They used the logarithmic scale in considering the Weber-Fechner law that indicates that the human eye distinguishes according to the logarithm of light intensity, although the validity of the application of Weber-Fechner law was unclear (Hebert, 1982).

Although a grading method is surely a cost-saving procedure for evaluating the pest abundance, we cannot currently analyze the results obtained by this method, since the meaning of averaging grade

* To whom correspondence should be addressed at: E-mail: yamamura@niaes.affrc.go.jp
is not clear. As an example, let us consider that the grades are 1 and 5 at two fields in one site, while they are 3 and 3 at two fields in another site. The averages of grades are 3 in both sites. However, it is not clear whether we can judge that the overall pest abundance in the two sites are the same or not. Thus, the grading method currently has only limited use. There are two questions that should be solved to construct a grading method that yields meaningful grades. (1) Which scale should we use to determine the boundary of grades? For example, is a logarithmic scale such as that divided at $x=1, 10, 100, 1,000$ preferable to an arithmetic scale such as that divided at $x=50, 100, 150, 200$? (2) Which grade width should we use? When we use a logarithmic scale, for example, is a 5-times width such as that divided at $x=1, 5, 25, 125$, preferable to a 10-times width such as that divided at $x=1, 10, 100, 1,000$? The purpose of this paper is to provide possible answers for these questions. We show that the average grade has a clear meaning if we construct the grade scale in accordance with the optimal transformation function that stabilize the variance. Then, we show the existence of an optimal width of grade that maximizes the accuracy of population estimates under a given labor. We can also show the existence of an optimal width that minimizes the labor under a given accuracy. The procedure to calculate the optimal width of grade is demonstrated by using hypothetical data.

**DETERMINATION OF GRADE SCALE**

A grading method is essentially a kind of non-parametric transformation. As an example, Fig. 1 shows the relation between the number of individuals, $x$, and the corresponding score in the Japanese official sampling procedure for the sampling survey of *Nephotettix cincticeps* cited above. We empirically adopt a large width of grade for a large number of individuals on an arithmetic scale in many cases. Hence, the grade score increases in nonlinear fashion with increasing $x$. In the case of the sampling survey of *N. cincticeps*, the score is close to $0.47(x+0.5)^{0.27}$ in a moderate range of $x$. Thus, the grading procedure for *N. cincticeps* is approximately equal to performing a power transformation with an exponent of 0.27.

One of the purposes of transformation is to enhance the homoscedasticity (the stability of error variance) that is required when we perform ANOVA (ANalysis Of Variance) to test the effect of factors on population abundance. The statistical power of ANOVA will be improved if homoscedasticity is improved by using a grading method. For example, Hollis Jr (1984) cited the following example: testing of herbicides against 15–20 weed species showed that percent evaluations of weed density yielded only 6–8 weed species that differentiate significantly between chemicals, whereas the Horsfall-Barratt system yielded significant differences in 12–18 species.

To determine an appropriate transformation formula, let us consider a situation that the population densities in two sites (sampling units) ($N_1$ and $N_2$) increase by a common factor $R$. After the increase, the population densities become $RN_1$ and $RN_2$, and hence the difference between the two population densities is given by $R(N_1-N_2)$. Thus, the difference is amplified by the factor $R$. In a natural logarithmic scale, in contrast, the difference is given by $\log_e(RN_1)-\log_e(RN_2)=\log_e(N_1)-\log_e(N_2)$. Thus, the difference is not influenced by the factor $R$. For this reason, the variance of population densities tends to become constant in a logarithmic scale. Therefore, a logarithmic transformation is recommended in this case. Partly due to the discrete nature of $x$, the transformation in the form of $\log_e(x)$ or $\log_{10}(x)$ cannot effectively stabilize the variance when $x$ is small. Yamamura (1999) suggested that the stabilization effect is improved if we add half of the discrete step to $x$. The discrete step is 1 in

![Fig. 1. Transformation effect implied in the grading method for estimating the abundance of *Nephotettix cincticeps* within the Japanese plant protection standards. Circles show the grade score allocated for the number of individuals, $x$. The curve shows a power transformation, $f(x)=0.47(x+0.5)^{0.27}$.](image-url)
the case of population counts, and hence the logarithmic transformation is modified to \( \log_{10}(x+0.5) \) or \( \log_{10}(x+0.5) \).

When there is dispersal across sites (sampling units), however, the above argument is somewhat violated. We frequently have the following relation whose parameters depend on the dispersal intensity of organisms (Yamamura, 2000):

\[
s^2 = am^b
\]

(1)

where \( m \) and \( s^2 \) are the mean and variance of the number of individuals in a sampling unit, and \( a \) and \( b \) are constants. This relation is called Taylor’s power law after Taylor (1961). The quantity of \( b \) equals 2 when there is no dispersal across sites, but decreases to approach 1 with increasing intensity of dispersal of organisms (Yamamura, 2000). We can derive an appropriate transformation formula, \( f(x) \), by using Eq. (1) (Yamamura, 1999):

\[
\begin{align*}
  f(x) &= (x+0.5)^{(b/2)} \quad \text{if } b \neq 2 \\
  f(x) &= \log_e(x+0.5) \quad \text{if } b = 2.
\end{align*}
\]

(2)

A logarithmic transformation is recommended by Eq. (2) if there is no dispersal across sites, as stated above. A square root transformation, \( f(x) = \sqrt{x+0.5} \), is recommended when \( b = 1 \), that is, when there is much dispersal across sites. Equation (2) is a power transformation with an exponent of \( 1 - (b/2) \). Hence, the grading method for \( N. cincticeps \) mentioned above is reasonable if the population of \( N. cincticeps \) follows Taylor’s power law with \( b = 2 \times (1 - 0.27) = 1.46 \). If we have sufficient data, we can determine an appropriate power transformation formula by substituting the estimate of \( b \) for Eq. (2). Equation (1) is described by a linear form in a logarithmic scale: \( \log_{10}(s^2) = \log_{10}(a) + b \log_{10}(m) \). Then, we can use a linear regression to estimate the parameters, since the variance of \( \log_{10}(s^2) \) is approximately constant (Perry, 1981).

For simplicity, however, a common logarithmic transformation, \( \log_{10}(x+0.5) \), will be recommended as the basis for constructing a grading method in many cases. Let \( w \) be the width of grade in a transformed scale. Then, a sampling unit is classified into the grade \( i \) if

\[
10^{w(i-1)} \leq (x+0.5) < 10^{wi}. \quad (i = 0, 1, 2, 3, \ldots)
\]

(3)

In other words, a sampling unit is classified into the grade \( i \) if

\[
10^{w(i-1)} \leq (x+0.5) < 10^{wi}. \quad (i = 0, 1, 2, 3, \ldots)
\]

(4)

For a grading method using \( 10^w = 5 \), for example, samples are classified into the following grades: \( x = 0, 1 \leq x \leq 4, 5 \leq x \leq 24, 25 \leq x \leq 124, 125 \leq x \leq 624, 625 \leq x \leq 3,124, \ldots \). The mid point of the minimum value and maximum value in a grade will be suitable as the score of the grade. For grade three in the above example, the lower and upper bounds are \( \log_{10}(25.5) \) and \( \log_{10}(124.5) \), respectively. Hence, \( [\log_{10}(25.5) + \log_{10}(124.5)]/2 = 1.751 \) is used as the score for grade three.

**DETERMINATION OF GRADE WIDTH**

The grade width influences the accuracy of estimates. The accuracy is usually evaluated by the mean square error defined by the expectation of the square of the difference between the estimate and the true value (Cochran, 1977). Let

\[
\begin{align*}
Y &= \text{transformed number of individuals}, \\
Z &= \text{grade score}, \\
z_j &= \text{observed grade score of the } j\text{th sample of a sampling survey}, \\
w &= \text{width of grade in the transformed scale}, \\
n &= \text{sample size, i.e., the number of sampling units examined in a sampling survey}, \\
E &= \text{operator indicating the expectation, and} \\
V &= \text{operator indicating the variance},
\end{align*}
\]

The population mean of the transformed number of individuals is estimated by the average of observed grade scores, \( \bar{z} = \frac{\sum_j z_j}{n} \). Then, the mean square error, which is denoted by \( MSE \), is given by

\[
\begin{align*}
MSE &= E[(\bar{z} - E(Y))^2] \\
&= E[(E(Z) - E(Y)) + (\bar{z} - E(Z))^2] \\
&= [E(Z) - E(Y)]^2 + \frac{V(Z)}{n},
\end{align*}
\]

(5)

Thus, we can calculate the mean square error by the sum of two quantities: the square of bias, \( [E(Z) - E(Y)]^2 \), and the variance of \( \bar{z} \), i.e., \( V(Z)/n \).

Figure 2 illustrates the relation between the frequency distribution under the exact counting method and that under a grading method. In a grading method, the probability of \( Y \) is combined within each grade to yield the probability of \( Z \). The
exact counting method corresponds to the limit case of $w \to 0$ in which the grade score is simply given by $Y$. With increasing width of grade, the square of bias, $[E(Z) - E(Y)]^2$, will increase because the discrepancy between the two distributions increases.

The mean square error is given by the sum of squared bias and variance as indicated by Eq. (5). The estimate of population density obtained from a grading method, $\tilde{z}$, has biases and hence we should reduce both the bias and the variance of estimate to obtain reliable estimates of population density. At this point, we are confronted with a dilemma. If we decrease the width of grade, we can decrease the bias of estimates. However, the time required to complete the survey of one sampling unit will simultaneously increase, given that we must classify the sampling unit into many grades. The total time available for sampling survey is limited. Hence, we are forced to decrease the total number of sampling units ($n$). Consequently, the variance of estimates simultaneously increases. If we increase the width of grade instead, the bias of estimates increases, but the variance of estimates simultaneously decreases because we can increase the total number of samples. These arguments indicate the existence of an optimal grade width that minimizes the mean square error of the estimate of population density, $\tilde{z}$.

In an ordinary two-stage sampling or a stratified random sampling, the optimal sample size is calculated by assuming that a fixed cost is required for examining each sampling unit irrespective of the number of individuals in the unit, although the actual cost will change depending on the number of individuals (Cochran, 1977; Adachi and Yamamura, 2000). Similarly, we first assume that a fixed time is required for examining a sampling unit under a given grade width irrespective of density. The problem accompanied with this assumption will be discussed later. Let $g(w)$ be the function expressing the time required for examining a sampling unit under a grade width $w > 0$. Then the total time required for the sampling survey, which is denoted by $C_{\text{total}}$, is given by

$$C_{\text{total}} = ng(w).$$

(6)

The quantity of $g(w)$ will decrease with increasing $w$. For simplicity, we use an exponential function to describe $g(w)$ approximately:

$$g(w) = C_0 \exp(-pw),$$

(7)

where $C_0$ is the time required for an exact counting method, and $p$ is a constant determining the rate of decrease of time with increasing width of grade. From Eqs. (6) and (7), we obtain

$$n = \frac{C_{\text{total}}}{C_0} \exp(pw).$$

(8)

For simplicity of explanation, we use the first order approximation to describe the square of bias, although it may be preferably described by a second order polynomial.

$$[E(Z) - E(Y)]^2 = dw,$$

(9)

where $d$ is a constant. The quantity of $V(Z)$ will fluctuate around $V(Y)$ in a complicated manner with increasing $w$. For simplicity of explanation, therefore, we treat $V(Z)$ as a constant. Then we denote

$$V(Z) = v,$$

(10)

where $v$ is a positive constant.

The optimal width of grade, which is denoted by $w^*$, is given by differentiating the mean square error with respect to $w$, and by letting it equal zero.

$$\frac{\partial \text{MSE}}{\partial w} \bigg|_{w=w^*} = 0.$$

(11)

When we use approximations given by Eqs. (7), (9), and (10), we obtain the following optimal grade width by substituting Eqs. (8), (9), and (10)
into Eq. (5) and by using the relation of Eq. (11).

$$w^* = \frac{1}{p} \log_e \left[ \frac{p \nu C_0}{d C_{\text{total}}} \right].$$  \hspace{1cm} (12)

The optimal sample size, which is denoted by $n^*$, is given by substituting $w^*$ for Eq. (8).

$$n^* = \frac{p \nu}{d},$$  \hspace{1cm} (13)

which is independent of $C_{\text{total}}$ and $C_0$. As intuitively expected, Eqs. (12) and (13) indicate that we should increase the sample size ($n$) by increasing the width of grade ($w$) if the parameter $\nu$ is large.

Optimal grade width to achieve a given accuracy ($MSE$) is also calculated in a similar way. By using Eqs. (5) and (6), we obtain

$$C_{\text{total}} = \frac{V(Z) g(w)}{MSE - \{E(Z) - E(Y)\}^2}.$$  \hspace{1cm} (14)

When we assume Eqs. (7), (9), and (10), we obtain

$$C_{\text{total}} = \frac{\nu C_0 \exp(-pw)}{MSE - \partial w}. \hspace{1cm} (15)$$

Then, we obtain the optimal width of grade by solving

$$\frac{\partial C_{\text{total}}}{\partial w} = 0.$$  \hspace{1cm} (16)

The solution is given by

$$w^* = \frac{MSE - \frac{1}{p}}{d},$$  \hspace{1cm} (17)

The quantity of $C_{\text{total}}$ is given by substituting $w^*$ for Eq. (15). Then, the number of sampling units is given by Eq. (8).

**NUMERICAL EXAMPLE**

Let us assume that we examined 10 sampling units and obtained the following number of insects in each unit: 107, 255, 19, 3, 38, 13, 6, 36, 25, and 164. We use only 10 sampling units for simplicity of explanation. More sampling units should be examined in actual situations to obtain reliable estimates of $E(Z)$, $V(Z)$, $E(Y)$, and $V(Y)$. It should be noted that such an intensive sampling procedure is required only once. If the optimal sampling plan is

<table>
<thead>
<tr>
<th>Score</th>
<th>Freq</th>
<th>Grade division</th>
<th>Score</th>
<th>Freq</th>
<th>Grade division</th>
<th>Score</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.544</td>
<td>1</td>
<td>0</td>
<td>-0.301</td>
<td>0</td>
<td>0</td>
<td>-0.301</td>
<td>0</td>
</tr>
<tr>
<td>0.813</td>
<td>1</td>
<td>1–4</td>
<td>0.415</td>
<td>1</td>
<td>1</td>
<td>0.577</td>
<td>2</td>
</tr>
<tr>
<td>1.130</td>
<td>1</td>
<td>5–24</td>
<td>1.065</td>
<td>3</td>
<td>10–99</td>
<td>1.510</td>
<td>5</td>
</tr>
<tr>
<td>1.290</td>
<td>1</td>
<td>25–124</td>
<td>1.751</td>
<td>4</td>
<td>100–999</td>
<td>2.501</td>
<td>3</td>
</tr>
<tr>
<td>1.407</td>
<td>1</td>
<td>125–624</td>
<td>2.447</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.562</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.585</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.031</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.216</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.407</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimate of mean score, $E(Z)$</th>
<th>1.499$^c$</th>
<th>1.551</th>
<th>1.621</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate of variance of score, $V(Z)$</td>
<td>0.356$^d$</td>
<td>0.418</td>
<td>0.507</td>
</tr>
<tr>
<td>Estimate of bias, $[E(Z) - E(Y)]$</td>
<td>—</td>
<td>0.052</td>
<td>0.122</td>
</tr>
<tr>
<td>Estimate of squared bias, $[E(Z) - E(Y)]^2$</td>
<td>—</td>
<td>$2.72 \times 10^{-3}$</td>
<td>$14.89 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

$^a$ Scores are simply given by $Y (= \log_{10}(x + 0.5))$ in this column.

$^b$ Scores are given by the mid point of the minimum and maximum value in the grade.

$^c$ This estimate corresponds to $E(Y)$.

$^d$ This estimate corresponds to $V(Y)$. 
once determined, we can estimate \( E(Y) \) with a given precision by using a grading method that requires the least labor as stated above.

To estimate the parameters of the model, \( d \) and \( v \), we re-classified the sampling units into grades. Two widths of grade, \( 10^w = 5 \) and \( 10 \), were used for this calculation. Upper part of Table 1 indicates the grade score and the frequency of sampling units that belong to each grade. The column of \( 10^w \rightarrow 1 \) indicates the exact counting method. Then, we calculated the average of score and the unbiased estimate of variance of score. The bias \( \hat{d} \) was estimated by the difference between average score of the grading method and that of the exact counting method. For example, for the grade width of \( 10^w = 5 \), we estimated the bias by \( 1.551 - 1.449 = 0.052 \). The squared biases were calculated by squaring these biases. The parameter \( d \) in Eq. (9) was estimated by the linear regression through origin, i.e., by the formula, \( \hat{d} = \frac{\sum [(\text{squared bias}) \times w]}{\sum w^2} \) (Snedecor and Cochran, 1989, p. 174). From Table 1, we obtained \( \hat{d} = [(2.72 \times 10^{-3}) \times \log_{10}(5) + (1.49 \times 10^{-3}) \times \log_{10}(10)]/[(\log_{10}(5)^2 + (\log_{10}(10))^2) = 0.0113 \). By averaging the estimates of \( \hat{V}(Z) \) of the exact counting (\( 10^w \rightarrow 0 \)) and two grade widths (\( 10^w = 5 \) and \( 10^w = 10 \)), we obtained an estimate of \( \hat{v} = (0.353 + 0.428 + 0.357)/3 = 0.427 \) from Table 1.

Equation (12) indicates that the optimal sampling plan is strongly influenced by the total time available for sampling survey (\( C_{\text{total}} \)). As a typical example, Fig. 3 shows the optimal grade width for \( \exp(-p) = 0.2 \), i.e., \( p = 1.61 \). The optimal width of grade decreases with increasing \( C_{\text{total}}/C_0 \). When \( (C_{\text{total}}/C_0) = 12 \), optimal grade width is \( 10^w = 10 \), whereas when \( (C_{\text{total}}/C_0) = 20 \), optimal grade width is \( 10^w = 5 \). As intuitively expected, an exact counting method, which corresponds to the limited case of \( 10^w = 1 \), becomes optimal in a case where \( C_{\text{total}} \) is sufficiently large.

**DISCUSSION**

Though it may seem likely that a rough examination of a sampling unit would yield only a rough estimate of density, this is not necessarily the case. A rough examination of each sampling unit enables us to increase the total number of samples, because the examination of a sampling unit can be completed quickly. Consequently, the increased number of samples may yield a superior estimate of density. The grading method discussed herein is a procedure that is potentially capable of enhancing the accuracy of estimates without increasing labor.

Previously, the average score of a grading method had been used as an index of infestation. However, because the meaning of the average score was unclear, it has rarely been used as an estimate of population density. Alternatively, several censoring methods have been used as time-saving methods for estimating the density of populations. For example, it is expedient to classify the samples into classes such as \( x = 0 \), \( x = 1 \), \( x = 2 \), and \( x \)\(\geq 3 \) by censoring the right-hand tail of distribution. Shiyomi (1974a, b) estimated the density of the individuals by using the maximum likelihood method, assuming that the distribution follows a negative binomial distribution. The estimation for this mode of censoring requires specific computer programs. A simpler form of censoring, by which the samples are classified into two classes, \( x = 0 \) and \( x \)\(\geq 1 \) seems to be more widely used. This kind of sampling, usually called a binomial sampling or a presence-absence sampling, has led to the development of various sampling methods. Two equations are principally used in this kind of estimation. The first, the Kono-Sugino equation, was empirically found by several authors independently (Kono and Sugino, 1958; Gerrard and Chiang, 1970; Nachman, 1984). The second, the Wilson-Room equation, uses a negative binomial distribution with Taylor’s (1961) power law (Wilson and Room, 1983). However, a presence-absence sampling is not applicable when the density of individuals is so high that most of
the sampling units contain at least one individual. Another tally threshold may be useful in such cases (Nyrop and Binns, 1991; Feng et al., 1993a,b). For example, the classification \( x < 3 \) and \( x \geq 3 \) may be sometimes useful. Conversely, such a large tally threshold is not effective when the density is so low that most of the sampling units are in the lower class. In contrast, compared to the presence-absence sampling, the grading methods discussed herein are effective at a wider range of densities.

We used a logarithmic scale in this paper, although other transformations may be preferable in some cases. If we use a square root transformation, we can apply the following boundary of grade in place of inequality (4):

\[
(wt)^2 \leq (x + 0.5) < [w(i+1)]^2. \quad (i=0,1,2,3,\ldots)
\]

If we use the estimated \( b \)-value, we obtain the following boundary from Eq. (2):

\[
(wt)^{2/(2-b)} \leq (x + 0.5) < [w(i+1)]^{2/(2-b)}. \quad (i=0,1,2,3,\ldots)
\]

However, it should be noted that a complicated scale may sometimes increase the time required for examining a sampling unit. Generally, it is preferable to use boundary values that observers can easily recognize. In a grading method of \( 10^v=5 \), although we used the boundary of \{1, 5, 25, 125, 625, \ldots\} in the calculation in the example (Table 1), it may be preferable to use boundary values such as \{1, 5, 25, 100, 500, \ldots\}, so that observers can recognize the boundary more easily.

Although we assumed that all sampling units are classified into correct grades, misclassification may occur in field observations, especially when the number of individuals is near the boundary of a grade. Such a misclassification alters the distribution of \( Z \) in Fig. 2, which affects the accuracy of the estimates of density. If the misclassification occurs in one direction, the bias of estimates will increase. For example, if sampling units are likely to be classified incorrectly into upper grades, the resultant estimate of density becomes larger. If the misclassification occurs in both directions (i.e., if sampling units are incorrectly classified into lower and upper grades by a comparable probability), the variance of estimates will increase. Thus, misclassification potentially increases both components of \( MSE \) in Eq. (5): the square of bias, \([E(Z)−E(Y)]^2\), and the variance of estimates, \( V(Z)/n \). If the influence of misclassification is suspected to be large, parameters such as \( d \) and \( v \) should be estimated from the distribution of grades that includes the degree of misclassification likely to occur in field observations. A degree of misclassification will be generally unavoidable, but we can reduce the possibility of misclassification by adopting a simpler boundary, as discussed above.

We derived the optimal sampling plan assuming a constant time is required to examine a sampling unit for simplicity. However, the time required for the sampling survey will change depending on the number of individuals in the sample in many cases. If the sample contains few or no individuals, we can readily complete the examination of the sample, whereas if the sample contains many individuals, we must spend a considerable amount of time for the sampling survey. Thus, the parameters \( C_0 \) and \( \rho \) in Eq. (7) will change depending on the density. Therefore, it is preferable to provide two or three sampling plans beforehand. For example, sampling plans for the situation of low and high densities. Then, by conducting a preliminary sampling, we can choose the most suitable sampling plan.

We discussed the simplest case for explanation in the above calculation; the squared bias linearly increases with increasing \( w \), and the variance \( V(Z) \) is approximately constant irrespective of \( w \). These simplifications may yield an incorrect result in several cases. It is preferable to use more flexible functions for describing these relations. The squared bias is preferably described by a second order polynomial with no intercept in several cases. The variance \( V(Z) \) is preferably described by a linear function of \( w \). The optimal sampling design is not given by an explicit form such as Eqs. (12) and (17) under these assumptions. However, we can derive the optimal sampling plan by iteratively minimizing \( MSE \) (Eq. (5)) or \( C_{total} \) (Eq. (14)).

ACKNOWLEDGEMENTS

We thank Dr. K. Kiritani for his helpful comments on the manuscript.

REFERENCES


