Optimal sample size for composite sampling with subsampling, when estimating proportion of pecky rice grains in a field

Kohji YAMAMURA and Masuhiro ISHIMOTO

Kohji Yamamura is Senior Researcher, Laboratory of Population Ecology, National Institute for Agro-Environmental Sciences, 3-1-3 Kannondai, Tsukuba 305-8604, Japan (E-mail: yamamura@niaes.affrc.go.jp). Masuhiro Ishimoto is Senior Researcher, Niigata Agricultural Research Institute, Crop Research Center, 857 Nagakura, Nagaoka, Niigata 940-0826, Japan.

The proportion of pecky rice grains has been empirically estimated using composite sampling with subsampling. The procedure is summarized as follows: (1) a fixed number of rice plants (n_1) are drawn at random in the paddy field; (2) all the rice grains in the collected rice plants are mixed well to form a composite; (3) a portion of the grains (n_2) are drawn at random from the composite; and (4) the collected grains are examined by eye to estimate the proportion of pecky rice grains. We propose a method to determine the optimal sample size in estimating the proportion of defective items by this kind of composite sampling with subsampling. Spatial heterogeneity in the proportion of defective items is included in the estimation. We use Taylor's power law to describe the density-dependent change of spatial heterogeneity. In controlling the precision of the estimate, we use the relative precision (D) that is defined by the coefficient of variation of the estimated proportion. We propose a rejection procedure where the product is rejected if the estimate of proportion with $D = 0.25$ is larger than a predetermined tolerable threshold of proportion. We further consider another control criterion where the consumer's risk (β) is controlled by a zero-tolerance method. The relation between two control criteria is examined.

Keywords: Consumer's risk; Increment sampling; Relative precision; Spatial heterogeneity; Taylor's power law

1. INTRODUCTION

The relative significance of insect pest species in paddy fields has greatly changed in the past 50 years in Japan (Kiritani 2006). The populations of the three classical major pests, the rice stem borer, *Chilo suppressalis*, the green rice leafhopper, *Nephotettix cincticeps*, and the small brown planthopper, *Laodelphax striatellus*, have greatly decreased in this time (Yamamura et al. 2006). The agricultural loss caused by these insects decreased accordingly. On the other hand, the agricultural loss caused by several rice bugs has increased since about 1995. The rice bugs are becoming the most serious insect pests in paddy fields. These rice bugs consist of several species of Pentatomidae, Coreidae, Alydidae and Miridae. The rice leaf bug, *Trigonotylus caelestialium* (Kirkaldy), is especially important in the northern part of Japan. These insects suck the rice grain to generate pecky rice grains. The existence of pecky rice grains causes serious problems even if the proportion of pecky rice grains is quite small. The grade of rice falls from first grade to second grade if the proportion of colored grains including pecky rice grains is larger than 0.001 (Ministry of Agriculture Forestry and Fisheries 2001). The market price of rice seriously decreases due to this downgrading. For example, the price of 30 kg rice grains may change from 7500 yen to 7000 yen.

The proportion of pecky rice grains is usually quite small, and hence a composite sampling procedure with subsampling has been empirically used in estimating the proportion. The procedure is summarized as follows: (1) a fixed number of rice plants are drawn at random in the paddy field and brought to the laboratory; (2) all the rice grains in the collected rice plants are mixed well to form a composite; (3) a portion of grains are drawn at random from the composite; and (4) the selected grains are examined by eye to estimate the proportion of pecky rice grains. A plant corresponds to the quantity called an increment. A rice grain corresponds to a sampling item.

Various composite sampling procedures that examine composites of sampling items instead of examining individual sampling items have been developed as cost-effective procedures (Boswell et al. 1988; Lovinson et al. 1994; Lancaster and Keller-McNulty 1998; United States Environmental Protection Agency 2000; Patil 2002). Some of these procedures are used for classification purposes, for example, to find HIV-infected persons effectively (Dorfman 1943; Sterrett 1957; Bhattacharyya et al. 1979; Emmanuel et al. 1988; Zenios and Wein 1998; Johnson and Patil 2001). Other procedures of composite sampling are used for estimation purposes. The parameters of interest are either continuous variables or binary variables. In applying the composite sampling to continuous variables, the theory has been developed for various sampling schemes that allow the subsampling from composites (Brown and Fisher 1972; Rohde 1976; Elder et al. 1980; Rohlf et al. 1996). In applying the composite sampling to binary variables, however, the application is mostly confined to the procedure called group testing (Chiang and Reeves 1962; Thompson 1962; Swallow 1985; Burrows 1987; Swallow 1987; Chen and Swallow 1990; Gastwirth and Johnson 1994; Hughes-Oliver and Swallow 1994; Chick 1996; Zenios and Wein 1998; Brookmeyer 1999; Colón et al. 2001; Tebbs et al. 2003; Hsu 2005; Bar-Lev et al. 2006; Yamamura and Hino 2007). In a group testing procedure for estimating the proportion of defective items, the sampling items are drawn at random and mixed, and all the items in the composite are used to examine the existence of defective items. This procedure is repeated several times to obtain an estimate of the proportion of defective items.

The sampling procedure for estimating the proportion of pecky rice grains is different from group testing procedures in two aspects: (1) rice grains are not drawn at random in the field, but rice grains are instead drawn as increments (i.e., clusters) given by rice plants; and (2) only a portion of the rice grains in the composite is measured by using subsampling from the composite. In this paper, we discuss the method of determining the sample size in estimating the proportion of defective items by this kind of sampling. In controlling the precision of the estimate, we use the relative precision (*D*) that is defined by the coefficient of variation (CV) of the estimated proportion. We use Taylor's power law to describe the relation between the mean and spatial heterogeneity in the proportion of defective items. We propose a rejection procedure where the product is rejected if the estimate of proportion with $D = 0.25$ is larger than a predetermined tolerable threshold of proportion. We further consider another control criterion where the consumer's risk (β) is controlled by a zero-tolerance method. The relation between two control criteria is examined.

2. SPECIFIC MODEL

We use the following notations.

s_i = the number of grains in the ith plant,

n_1 = the number of drawn rice plants,

n_2 = the number of rice grains drawn from the composite ($n_2 \leq s_i n_1$),

P_i = the probability that a rice grain around the ith plant is pecky,

P_0 = the average of P_i over the sampling field, i.e., $P_0 = E(P_i)$,

X_i = the number of pecky rice grains in the ith plant,

Y = the total number of pecky rice grains in the composite, i.e., $Y = \sum_1^{n_1} X_i$,

Z = the number of pecky rice grains in the n_2 grains that are drawn from the composite.

We want to estimate P_0, that is, the average of the proportion of pecky rice grains. The proportion of pecky rice grains per plant fluctuates depending on the spatial position of the rice plants in the field. We assume that the proportion of pecky rice grains fluctuate following a distribution. This type of estimation is sometimes called 'model-based approach' (Lohr 1999, p47; Thompson 2002, p22). For simplicity, we first assume a specific form of distribution in describing the spatial heterogeneity. A beta distribution is frequently used to describe the heterogeneity in the proportions. The proportion of pecky rice grains is usually very small. A beta distribution can be approximately described by a gamma distribution when the average proportion is small, and hence we use a gamma distribution that is more tractable than a beta distribution. The probability density of P_i is given by

$$f(p) = \frac{1}{\Gamma(k)} \lambda^k p^{k-1} \exp(-\lambda p), \qquad (2.1)$$

where k and λ are the shape parameter and scale parameter, respectively. The mean (P_0) and variance ($V(P_i)$) are given by k/λ and k/λ^2, respectively. Thus, we are estimating the parameter of the model, k/λ, in this case. The number of pecky rice grains in the ith plant (X_i) for a given P_i is given by a binomial distribution, but we can use a Poisson distribution as an approximation if the P_i is sufficiently small. If we use a distribution which is conditional to s_i, we have

$$\Pr(X_i = x \mid P_i = p) = \frac{1}{x!}(s_i p)^x \exp(-s_i p), \qquad (2.2)$$

Then, the probability density of X_i is given by

$$\Pr(X_i = x) = \int_0^\infty \Pr(X_i = x \mid P_i = p) f(p) dp$$
$$= \frac{\Gamma(k+x)}{x!\Gamma(k)} \left(1 + \frac{s_i}{\lambda}\right)^{-k} \left(\frac{s_i}{\lambda + s_i}\right)^x, \qquad (2.3)$$

which is a negative binomial distribution with mean $s_i k/\lambda$ and variance $s_i k(\lambda + s_i)/\lambda^2$.

For simplicity, we assume that the number of grains in a plant (s_i) is almost the same for all plants, and we denote it by s by omitting the subscript. We assume that s is known to us beforehand. Then, the distribution of Y is given by the n_1 times convolution of (2.3). We readily obtain the following distribution due to the reproducibility of the negative binomial distribution (Minotani 2003).

$$\Pr(Y = y) = \frac{\Gamma(n_1 k + y)}{y!\Gamma(n_1 k)} \left(1 + \frac{s}{\lambda}\right)^{-n_1 k} \left(\frac{s}{\lambda + s}\right)^{y}. \qquad (2.4)$$

The mean and variance of Y are given by sn_1k/λ and $sn_1k(\lambda + s)/\lambda^2$, respectively.

The probability distribution of the number of pecky rice grains (Z) in the n_2 grains that are obtained by a subsampling from the composite is most exactly expressed by a hypergeometric distribution.

$$\Pr(Z = z \mid Y = y) = \binom{y}{z}\binom{n_1 s - y}{n_2 - z} \Big/ \binom{n_1 s}{n_2}. \qquad (2.5)$$

If the proportion of pecky rice grains in the composite ($Y/(n_1 s)$) is sufficiently small, we can approximately consider that each pecky rice grain in the composite is drawn by an equal probability, $n_2/(n_1 s)$. Kuno (1991) used this form of binomial approximation of hypergeometric distribution in deriving the sample size for zero-tolerance method.

$$\Pr(Z = z \mid Y = y) = \binom{y}{z} \left(\frac{n_2}{n_1 s}\right)^{z} \left(1 - \frac{n_2}{n_1 s}\right)^{y - z}. \qquad (2.6)$$

Then, the number of pecky rice grains (Z) in the subsample is given either by a compound distribution (i.e., a stopped-sum distribution by the terminology of Johnson et al. 2005) or a mixture distribution; a compound negative binomial distribution that is compounded with Bernoulli distribution of the parameter $n_2/(n_1 s)$; or a mixture binomial distribution where the number of trials (Y) follows a negative binomial distribution. The compound distribution or the mixture distribution yields the negative binomial distribution where the mean is multiplied by $n_2/(n_1 s)$ while the shape parameter is the same (Shimizu 2006).

$$\Pr(Z = z) = \frac{\Gamma(n_1 k + z)}{z!\Gamma(n_1 k)} \left(1 + \frac{n_2}{n_1 \lambda}\right)^{-n_1 k} \left(\frac{n_2}{n_1 \lambda + n_2}\right)^{z}. \tag{2.7}$$

The mean and variance of Z are given by $n_2 k/\lambda$ and $n_2 k(n_1\lambda + n_2)/(n_1\lambda^2)$, respectively. The estimate of P_0 is given by

$$\hat{P}_0 = Z/n_2. \tag{2.8}$$

Hence, (2.7) yields the following mean and variance of the estimate.

$$E\left(\hat{P}_0\right) = k/\lambda = P_0, \tag{2.9}$$

$$V\left(\hat{P}_0\right) = \frac{k(n_1\lambda + n_2)}{n_1 n_2 \lambda^2}. \tag{2.10}$$

3. GENERAL MODEL

We can extend the above argument to general distributions of the proportion of pecky rice grains. The mean and variance of pecky rice grains (X_i) in the ith plant with a known s_i are given by

$$E\left(X_i\right) = \int_0^\infty E(X_i \mid P_i = p) f(p)dp$$

$$= \int_0^\infty s_i p f(p)dp = s_i P_0, \tag{3.1}$$

$$V\left(X_i\right) = \int_0^\infty E((X_i - s_i P_0)^2 \mid P_i = p) f(p)dp$$

$$= \int_0^\infty (E(X_i - s_i P_i)^2 + (s_i P_i - s_i P_0)^2 \mid P_i = p) f(p)dp$$

$$= \int_0^{\infty} (s_i p + s_i^2 (p - P_0)^2) f(p) dp$$
$$= s_i P_0 + s_i^2 V(P_i) . \qquad (3.2)$$

The rearrangement of (3.1) and (3.2) yields the following equation.

$$P_0 = E\left(\frac{X_i}{s_i}\right), \qquad (3.3)$$
$$V(P_i) = V\left(\frac{X_i}{s_i}\right) - E\left(\frac{X_i}{s_i^2}\right). \qquad (3.4)$$

We can obtain the moment estimates of P_0 and $V(P_i)$ by substituting the mean and variance of X_i/s_i and the mean of X_i/s_i^2 into (3.3) and (3.4). Numerical examples of calculation are shown in the electronic material that is stored in JABES site (http://www.amstat.org/publications/jabes/data.shtml). As for the mean and variance of the total number of pecky rice grains in the composite (Y), we have the following equation.

$$E(Y) = \sum_{i=1}^{n_1} s_i P_0 , \qquad (3.5)$$
$$V(Y) = \sum_{i=1}^{n_1} (s_i P_0 + s_i^2 V(P_i)) . \qquad (3.6)$$

For simplicity, we again assume that the number of grains in a plant (s_i) is almost the same for all plants, and we denote it by s by omitting the subscript. We assume that s is known to us beforehand.

The probability distribution of the number of pecky rice grains (Z) in the n_2 grains that are obtained by a subsampling from the composite is expressed by a hypergeometric distribution given by (2.5). The mean and variance are given by

$$E(Z \mid Y = y) = \frac{n_2 y}{n_1 s}, \tag{3.7}$$

$$\begin{aligned} V(Z \mid Y = y) &= \frac{n_2(n_1 s - n_2)}{n_1 s - 1} \frac{y}{n_1 s}\left(1 - \frac{y}{n_1 s}\right) \\ &\approx \frac{n_2(n_1 s - n_2)}{n_1 s} \frac{y}{n_1 s} \\ &= y \frac{n_2}{n_1 s}\left(1 - \frac{n_2}{n_1 s}\right). \end{aligned} \tag{3.8}$$

We again used a binomial approximation where y is the number of trials and $n_2/(n_1 s)$ is the probability of occurrence. This approximation will be valid because we are considering a situation where the proportion of pecky rice grains in the composite ($Y/(n_1 s)$) is sufficiently small. The expectation of $\hat{P}_0$ is given by

$$\begin{aligned} E\left(\hat{P}_0\right) &= \frac{1}{n_2} \sum_{y=0}^{\infty} \sum_{z=0}^{y} z \Pr(Z = z \mid Y = y) \Pr(Y = y) \\ &= P_0 . \end{aligned} \tag{3.9}$$

Thus, $\hat{P}_0$ is an unbiased estimate of P_0. The variance of $\hat{P}_0$ is given by

$$\begin{aligned} V\left(\hat{P}_0\right) &= \frac{1}{n_2^2} \sum_{y=0}^{\infty} \sum_{z=0}^{y} \left[z - E\left(Z\right)\right]^2 \Pr(Z = z \mid Y = y) \Pr(Y = y) \\ &= \frac{1}{n_2^2} \sum_{y=0}^{\infty} \sum_{z=0}^{y} \left[\left(z - E(Z \mid Y = y)\right)^2 + \left(E(Z \mid Y = y) - E\left(Z\right)\right)^2\right] \Pr(Z = z \mid Y = y) \Pr(Y = y) \\ &\approx \frac{1}{n_2^2} \sum_{y=0}^{\infty} \left[y \frac{n_2}{n_1 s}\left(1 - \frac{n_2}{n_1 s}\right)\right] \Pr(Y = y) + \frac{1}{n_2^2}\left(\frac{n_2}{n_1 s}\right)^2 V(Y) \\ &= \frac{P_0}{n_2} + \frac{V(P_i)}{n_1}, \end{aligned} \tag{3.10}$$

where we used the binomial approximation of (3.8). The rearrangement of (3.9) and (3.10) yields the following equation.

$$V(P_i) = n_1 V(\hat{P}_0) - \left(\frac{n_1}{n_2} \right) E(\hat{P}_0) . \quad (3.11)$$

If we could estimate the mean and variance of $\hat{P}_0$ by repeating the composite sampling in the same field, we can obtain the moment estimates of P_0 and $V(P_i)$ by using (3.9) and (3.11), although such an estimation is not practical in our case.

4. DEFINITION OF PRECISION

If the distribution of $\hat{P}_0$ could be approximately described by a normal distribution with a mean P_0 and a fixed variance $V(\hat{P}_0)$, the 95% confidence interval of $\hat{P}_0$ is given by $\hat{P}_0 \pm 1.96\sqrt{V(\hat{P}_0)}$. We can keep the width of confidence interval at a constant by controlling $V(\hat{P}_0)$ at a constant. However, P_0 does not have negative quantity; P_0 will be rather determined by a multiplicative manner than by an additive manner. Hence, it will be preferable to keep the relative width of confidence interval at a constant instead of keeping the absolute width of confidence interval at a constant. Then, instead of controlling the quantity of $V(\hat{P}_0)$ at a constant, we control the precision by the relative precision (D) defined by the coefficient of variation (CV) of the estimates (Kuno 1986):

$$D = \sqrt{V(\hat{P}_0)} \Big/ P_0 . \quad (4.1)$$

The required quantity of D changes depending on the purpose of the control. By substituting (3.10) into (4.1), we obtain

$$D^2 = \frac{1}{n_2 P_0} + \frac{V(P_i)}{n_1 P_0{}^2}. \quad (4.2)$$

For a fixed set of n_1 and D, we can calculate the required quantity of n_2 by

$$n_2 = \frac{n_1 P_0}{D^2 n_1 P_0{}^2 - V(P_i)}. \quad (4.3)$$

We have an inequality, $sn_1 \geq n_2$. Hence, (4.3) indicates that n_1 must satisfy the following condition:

$$n_1 \geq \frac{P_0 + sV(P_i)}{D^2 s P_0{}^2}. \quad (4.4)$$

5. TAYLOR'S POWER LAW

The spatial distribution of the proportion of pecky rice grains will change depending on the spatial distribution of rice bugs. The distribution of insects can be generally described by Taylor's power law (Taylor 1961; Taylor et al. 1978; Taylor et al. 1979; Taylor 1984). Let μ and σ^2 be the spatial mean and variance of the number of insects, respectively. Then, we frequently find a linear relation called Taylor's power law:

$$\log_e(\sigma^2) = \log_e(a) + b\log_e(\mu), \quad (5.1)$$

where a and b are constants. The parameter b usually lies between 1 and 2. This linear relation is approximately generated if the density of individuals increases with increasing occupied area. A nearly exact linear relation emerges when the quantity of (instantaneous increasing rate of density)/(instantaneous expansion rate of occupied area) is constant (Yamamura 2000). Taylor's power law will be also applicable to the variance of the distribution of the proportion of pecky rice grains because the proportion of pecky rice grains will be approximately proportional to the incidence of rice bugs.

$$\log_e[V(P_i)] = \log_e(a) + b\log_e(P_0)\,. \tag{5.2}$$

We obtain the following equations by substituting (5.2) into (4.2) and (4.3).

$$D^2 = \frac{1}{n_2 P_0} + \frac{a P_0^{\,b-2}}{n_1}, \tag{5.3}$$

$$n_2 = \frac{n_1}{D^2 n_1 P_0 - a P_0^{\,b-1}}\,. \tag{5.4}$$

The range of n_1 given by (4.4) is expressed by

$$n_1 \geq \frac{s a P_0^{\,b-1} + 1}{D^2 s P_0}\,. \tag{5.5}$$

6. CONTROL OF PRECISION

We can obtain the combination of n_1 and n_2 that achieves the relative precision D by using (5.4). However, we have a contradiction in this respect; we must know the quantity of P_0 beforehand to estimate P_0 with a specified precision by using (5.4). We can avoid this contradiction by determining the critical quantity of P_0 that is tolerable. Let us denote the critical quantity by P_c. The derivative of (5.3), $\partial(D^2)/\partial P_0$, is negative at least if $b < 2$. Therefore, if we use a set of n_1 and n_2 that achieves a given relative precision (D) under P_c, we can achieve a smaller D for all P_0 values that are larger than P_c. Thus, the precision of the estimates becomes superior for a larger P_0; we can enhance the precision of estimates for a worse case than P_c. In that sense, we can control the precision by using the set of n_1 and n_2 that are calculated at the critical proportion P_c.

We can calculate the optimal combination of n_1 and n_2 that minimizes the total cost of inspection. Let c_1 be the cost that is required to collect one rice plant, and let c_2 be the cost that is required to examine one rice grain. Then, the total cost (c_{total}) is given by

$$c_{total} = c_1 n_1 + c_2 n_2. \quad (6.1)$$

Let n_1* and n_2* be the optimal quantities of n_1 and n_2, respectively, that minimize the total cost to achieve a fixed amount of D. We can calculate n_1^* by solving $\partial c_{total} / \partial n_1 = 0$ after substituting (5.4) into (6.1) to eliminate n_2.

$$n_1^* = \sqrt{\frac{c_2}{c_1} \frac{\sqrt{a}}{D^2} P_c^{0.5b-1.5}} + \frac{a}{D^2} P_c^{b-2}. \quad (6.2)$$

The corresponding quantity of n_2^* is calculated by using (5.4). Actual sample size is an integer, and hence the exact optimal quantity of n_1 is either integer just below n_1^* or just above n_1^*.

To minimize the quantity D for a fixed total cost (c_{total}), let n_1' and n_2' be the optimal quantities of n_1 and n_2, respectively. Then n_1' is found to be

$$n_1' = \frac{c_{total}\sqrt{aP_c^{b-1}}}{\sqrt{c_1}\left(\sqrt{c_2} + \sqrt{ac_1 P_c^{b-1}}\right)}. \quad (6.3)$$

The corresponding quantity of n_2' is calculated by using (6.1).

It should be noted that the cost function of (6.1) is different from that of a two-stage sampling. The total cost in (6.1) is more precisely described by

$$c_{\text{total}} = c_0 + c_1 n_1 + c_2 n_2, \quad (6.4)$$

where c_0 is the cost for making a composite. If we adopt a two-stage sampling in which grains are drawn separately from each sampled plant, we can estimate the proportion of pecky rice grains by using the known formula for two-stage sampling for proportion (Cochran 1977, p279). In a two-stage sampling, the total cost is roughly given by

$$c_{\text{total}} = (c_0 + c_1) n_1 + c_2 n_2, \quad (6.5)$$

where the cost for preparing composites becomes n_1 times larger than that of a composite sampling. Hence, a two-stage sampling is scarcely adopted for the estimation of the proportion of pecky rice grains except for experimental purposes.

7. REJECTION PROCEDURE USING D = 0.25

In judging the difference in the abundance of insect pests, half of the quantity of the density is frequently used empirically in Japan. In the official experiments used to evaluate the effectiveness of pesticides, for example, the abundance of insects in treated fields is judged to be smaller than that in control fields if the observed difference is larger than 50% of the quantity of the abundance in control fields (Japan Plant Protection Association 2003). The definition in (4.1) indicates that the standard error (SE) is given by DP_0. The 95% confidence limits are given by ±1.96×SE in a case of a normal distribution with a fixed variance. If we use D = 0.25, therefore, the 95% confidence interval of $\hat{P}_0$ is approximately given by $[0.5\hat{P}_0, 1.5\hat{P}_0]$ under the normal approximation with a fixed variance. Hence, we empirically use $D = 0.25$ as a standard, although $\hat{P}_0$ will not exactly follow a normal distribution.

In the actual inspection of the proportion of pecky rice grains, we must judge the acceptance of product by using $\hat{P}_0$. We consider a rejection procedure where we reject the product of a field if we obtain $\hat{P}_0 > P_c$. The probability of acceptance is about 0.975 at $P_0 = 0.5P_c$, 0.5 at $P_0 = P_c$, and 0.025 at $P_0 = 1.5P_c$, if the distribution of $\hat{P}_0$ follows a normal distribution with a fixed variance, under the sampling procedure that controls the relative precision at $D = 0.25$. Therefore, the producer's risk (i.e., the probability of false rejection of a superior product) is 0.025 at $P_0 = 0.5P_c$. The consumer's risk (i.e., the probability of false acceptance of an inferior product) is 0.025 at $P_0 = 1.5P_c$. Actual distribution of $\hat{P}_0$ will not follow a normal distribution with a fixed variance. However, we will be able to expect that a superior product having $P_0 < 0.5P_c$ is almost certainly accepted while an inferior product having $P_0 > 1.5P_c$ is almost certainly rejected.

8. CONTROL OF CONSUMER'S RISK

We can exactly control the consumer's risk (that is denoted by β), if we assume a specified distribution for the distribution of P_i. Let us consider a sampling inspection where the product of a field is rejected if the final sample contains at least one pecky rice grain. We call this rejection procedure as 'zero-tolerance method'. The zero-tolerance method is not adopted in actual examination of pecky rice grains, but it is used in various inspections such as the import plant quarantine inspection in Japan (Yamamura and Sugimoto 1995), because it requires the smallest sample size. The sample size (n) in the Japanese official procedure of import plant quarantine inspection was principally calculated by using the following formula that is based on the Poisson distribution.

$$n = -\frac{\log_e(\beta)}{P_c}. \quad (8.1)$$

In most of the standards in industrial inspection such as international standards (ISO) and Japanese industrial standards (JIS), the consumer's risk is set at $\beta = 0.1$ while the producer's risk (that is usually denoted by α) is set at 0.05. However, in the Japanese import plant quarantine inspection, the consumer's risk was set at $\beta = 0.05$ in principle, because the primary purpose of the quarantine inspection lies in the protection of consumer's risk rather than producer's risk. If we use $\beta = 0.05$, (8.1) is approximately given by $3/P_c$, and hence we can easily calculate the sample size by a mental calculation. This rule is sometimes called the 'rule of three' (Jovanovic and Levy 1997; van Belle 2002; Iwasaki 2005).

The probability that no pecky rice grains are included in the final sample in the composite sampling is given by the zero-term of (2.7) if we assume a gamma distribution for P_i. We calculate the sample size so that we have $\Pr(Z = 0) < \beta$. By equating the mean and variance of (2.1) to P_0 and aP_0^b, respectively, we obtain $k = P_0^{2-b}/a$ and $\lambda = P_0^{1-b}/a$. By substituting $k = P_0^{2-b}/a$ and $\lambda = P_0^{1-b}/a$ into the zero-term of (2.7), we obtain

$$\left(1 + \frac{n_2 a P_0^{b-1}}{n_1}\right)^{-\frac{n_1}{aP_0^{b-2}}} < \beta. \quad (8.2)$$

The first derivative of the left hand side of (8.2) about P_0 is negative at least if $1 < b < 2$. Therefore, we can obtain the combination of n_1 and n_2 to achieve the above inequality for all P_0 satisfying $P_c < P_0 \le 1$, by finding the quantities that satisfy $\Pr(Z = 0 | P_0 = P_c) = \beta$. Hence, we have the following relation.

$$n_2 = \frac{n_1}{aP_c^{\,b-1}}\left(\beta^{-\frac{aP_c^{\,b-2}}{n_1}} - 1\right). \tag{8.3}$$

If the heterogeneity $V(P_i)$ approaches zero while P_0 is kept at a constant in (2.1), the quantity of λ increases to infinity while k/λ is kept at a constant. Then, the zero-term of (2.7) becomes $\exp(-n_2 P_0)$. Therefore, (8.3) reduces to (8.1) where n is replaced by n_2. If $c_1 > 0$, the corresponding optimal quantity of n_1 is given by the smallest integer that satisfies $sn_1 \geq n_2$, because n_1 does not influence the consumer's risk if there is no heterogeneity.

Let $n_1^{\#}$ and $n_2^{\#}$ be the optimal quantities of n_1 and n_2, respectively, that minimize the total cost to achieve a fixed amount of β. We can calculate $n_1^{\#}$ by solving $\partial c_{\text{total}} / \partial n_1 = 0$ after substituting (8.3) into (6.1) to eliminate n_2.

$$n_1^{\#} = -\frac{aP_c^{\,b-2}\log_e(\beta)}{1 + W\left(\frac{a(c_1/c_2)P_c^{\,b-1} - 1}{\exp(1)}\right)}, \tag{8.4}$$

where W is the Lambert W-function. The corresponding quantity of $n_2^{\#}$ is calculated by using (8.3). Actual sample size is an integer, and hence the exact optimal quantity of n_1 is either integer just below $n_1^{\#}$ or just above $n_1^{\#}$.

9. APPLICATION

We must first estimate the parameters of Taylor's power law (a and b) for the proportion of pecky rice grains. We collected data from 16 paddy fields (each area ranges from 75 to 150 m^2) in Niigata Agricultural Research Institute, Crop Research Center (37°

26' N, 138° 52' E). The rice leaf bug, *Trigonotylus caelestialium*, is the primary species causing pecky rice grains in these fields. Three varieties were used: Wasejiman (5 fields), Koshi-ibuki (5 fields), and Koshi-hikari (6 fields). We selected 20 plants at random from each field. We recorded the number of pecky rice grains (X_i) and the total number of grains (s_i) separately for each plant. The data used in this estimation are stored in the JABES web-site (http://www.amstat.org/publications/jabes/data.shtml). We can estimate the mean and variance for each field by the moment method by using (3.3) and (3.4). However, we instead obtained maximum likelihood estimates, that will be more reliable, by maximizing the sum of the logarithm of (2.3). The procedure given in Appendix A of Yamamura and Sugimoto (1995) was used for the estimation. We then obtained 16 pairs of estimates, $\hat{k}$ and $\hat{\lambda}$. The quantities of P_0 and $V(P_i)$ were estimated by $\hat{k}/\hat{\lambda}$ and $\hat{k}/\hat{\lambda}^2$, respectively. Then, we estimated the power law relation from the 16 pairs of $\hat{P}_0$ and $\hat{V}(P_i)$ by using a linear regression of the form of (5.2) (Figure 1). The explanatory variable, $\hat{P}_0$ in this case, has some measurement errors, but a linear regression will suffice in estimating the parameters of power law (Perry 1981). The estimated parameters (±SE) were $\widehat{\log(a)} = -2.19 \pm 0.81$ and $\hat{b} = 1.60 \pm 0.13$. If we used the moment method in estimating P_0 and $V(P_i)$, the estimated parameters (±SE) were $\widehat{\log(a)} = -2.41 \pm 0.66$ and $\hat{b} = 1.58 \pm 0.11$ (See the electronic material that is stored in JABES site). The curves in Figure 2 show the combination of n_1 and n_2 that satisfy $D < 0.25$ for all P_0 in a range of $P_0 > P_c$. The combination is shown for several quantities of P_c. The number of grains in a plant was set at 1400, which seems to be the average number of grains per plant. The required n_2 decreases in a decelerating manner with increasing n_1.

Figure 1

Figure 2

We must estimate the costs (c_1 and c_2), to determine the optimal sample size. About 60 seconds are required in drawing a rice plant and in shelling the rice grains. About 0.12 seconds are required to examine a rice grain on average. We thus use $c_1/c_2 = 60/0.12 = 500$. The grade of rice falls from the first grade to the second grade if the proportion of pecky rice grains is larger than 0.001 (Ministry of Agriculture Forestry and

Fisheries 2001). Thus, we use $P_c = 0.001$. Equation (6.2) indicates that the optimal n_1 is 58 in this case. Then, (5.4) indicates that the corresponding n_2 is about 31000. We can alternatively obtain the optimal combination of n_1 and n_2 in a graphical manner. We have $n_2 = (c_{total}/c_2) - (c_1/c_2)n_1$ from (6.1). In our case, we have $n_2 = (c_{total}/c_2) - 500n_1$. The minimization of c_{total} is equivalent to the minimization of the intercept, c_{total}/c_2. Hence, we can visually find the optimal combination of n_1 and n_2 for $P_c = 0.001$ by finding the point where the curve of $P_c = 0.001$ contacts a line with a slope of −500 (see Figure 2).

If we explicitly assume a gamma distribution for the distribution of P_i, we can use (2.7) to obtain the 'exact' confidence interval that is defined by the inverse of testing. By equating the mean and variance of (2.1) to P_0 and $aP_0^{\,b}$, respectively, we obtain $k = P_0^{\,2-b}/a$ and $\lambda = P_0^{\,1-b}/a$. After substituting the parameters $k = P_0^{\,2-b}/a$, $\lambda = P_0^{\,1-b}/a$, $n_1 = 58$, and $n_2 =$ 31000 into (2.7), we obtain the lower 95% exact confidence limit by numerically finding P_0 where the upper 2.5% point coincides to $\hat{P}_0$. Similarly, we obtain the upper 95% exact confidence limit by numerically finding P_0 where the lower 2.5% point coincides to $\hat{P}_0$. For example, if we obtain an estimate $\hat{P}_0 = 0.001$, i.e., $Z = 31$, the exact confidence interval is [0.00060, 0.00164]. This exact confidence interval is close to the confidence interval that is calculated by the normal approximation, $\hat{P}_0 \pm 1.96 D\hat{P}_0$, that is [0.00051, 0.00149]. Thus, the normal approximation seems satisfactory in this case.

We can calculate another optimal sample size by controlling the consumer's risk if we assume a gamma distribution for the distribution of P_i. By substituting the cost ratio ($c_1/c_2 = 500$) into (8.4) and (8.3), we estimate the optimal size of n_1 and n_2 to be $n_1^{\#} = 5.5$ and $n_2^{\#} = 5001$, respectively (Figure 3). The exact optimal size of n_1 and n_2 were 6 and 4775, respectively. Equation (8.1) indicates that the required sample size is 2996 if we perform a simple random sampling. Thus, the required number of grains increases from 2996 to 4775 if we use a composite sampling with subsampling instead of using a simple random sampling, although we can greatly reduce the cost required for sampling by adopting composite sampling.

Figure 3

The optimal sample size for controlling the consumer's risk at $\beta = 0.05$ by a zero-tolerance method (i.e., $n_1^{\#} = 5.5$ and $n_2^{\#} = 5001$) was much different from the optimal sample size for controlling the relative precision at $D = 0.25$ (i.e., $n_1^* = 58$ and $n_2^* = 31000$). The zero-tolerance method requires a much smaller cost for inspection, but we must tolerate a larger producer's risk if we use such a small sample size. The conceptual difference between two rejection procedures is shown in Figure 4. The broken curve in Figure 4 indicates the operating characteristic curve (OC-curve) for the zero-tolerance method using $P_c = 0.001$ and $\beta = 0.05$ under the assumption that P_i fluctuates following a gamma distribution. The probability of acceptance is exactly 0.05 at $P_0 = 0.001$ by its definition. However, the probability of acceptance is small even if P_0 is much smaller than P_c; for example, the probability of acceptance is 0.18 even if $P_0 = 0.0005$. Let us next consider the rejection procedure where we reject the product if we obtain $\hat{P}_0 > P_c$ under the sampling procedure that controls $D = 0.25$ for $P_c = 0.001$. The solid curve in Figure 4 indicates that the producer's risk is much smaller in this rejection procedure under the assumption that P_i fluctuates following a gamma distribution. The probabilities of acceptance at $P_0 = 0.0005$, 0.001, and 0.0015 are 0.995, 0.502, and 0.046, respectively. As we expected for general distributions, a superior product having $P_0 < 0.5P_c$ is almost certainly accepted while an inferior product having $P_0 > 1.5P_c$ is almost certainly rejected if we use $D = 0.25$.

Figure 4

10. DISCUSSION

We proposed a method to find the optimal sample size for the composite sampling with subsampling. This method will be applicable to various sampling procedures that have the following characteristics. (1) The purpose of sampling is the estimation of the proportion of defective items in a lot. In our case, a defective item corresponds to a pecky rice grain; a lot corresponds to a paddy field. (2) The proportion of defective items is sufficiently small. (3) The sampling items are drawn by increments, that is, several items are drawn at the same position. In our case, an increment corresponds to a plant. (4) All collected items are mixed well to yield a composite. (5) A portion of the collected items are examined by performing a subsampling from the composite.

Composite sampling with subsampling has been widely used in the inspection of agricultural products. In the procedure to estimate the proportion of commingling of genetically modified organisms (GMOs) of corn, for example, sampling items are drawn by increments, and the increments are mixed and subsequently subsampled to form 'laboratory sample' (Japanese Ministry of Health Labor and Welfare 2001). Similar sampling procedures are used in several international standards (ISO 1990, 1999, 2000, 2002, 2003) and FGIS (Federal Grain Inspection Service 1995). When these types of sampling are used for the estimation of the proportion of defective items, the required sample size is sometimes calculated by assuming a simple random sampling. Such a sample size is valid only if the defective items are distributed at random in the lot, that is, only if there is no spatial heterogeneity in P_i. The variance (3.10) becomes P_0/n_2 if $V(P_i) = 0$; that is, the variance becomes identical to the variance of the average of n_2 Poisson variables of mean P_0. However, if the distribution of defective items is highly aggregated in the lot, the variance will be severely underestimated if we use the formula for a simple random sampling. We should use appropriate equations such as (3.10) if large heterogeneity is suspected. The corresponding optimal sample size should be calculated by using (6.2) and (5.4). Similarly, if we perform zero-tolerance method, we should use (8.4) and (8.3) instead of (8.1) in determining the sample size if large heterogeneity is suspected.

In constructing the optimal sampling design, we should first estimate the parameters of Taylor's power law unless we have clear evidence that the variance $V(P_i)$ is constant irrespective of P_0. If we examine the number of defective items of each increment separately, we can estimate P_0 and $V(P_i)$ of each lot by using the moment method by using (3.3) and (3.4). Alternatively, if we can repeat composite sampling several times in each lot, we obtain $\hat{E}(\hat{P}_0)$ and $\hat{V}(\hat{P}_0)$ for each lot. Then, we can estimate P_0 and $V(P_i)$ of each lot by the moment method by using (3.9) and (3.11). We can estimate the parameters of Taylor's power law by using a linear regression such as that in Figure 1 by using several pairs of the estimates of P_0 and $V(P_i)$. If we have a clear evidence that the variance $V(P_i)$ is constant irrespective of P_0, only a single estimate of $V(P_i)$ is required; we can obtain the optimal sample size by substituting $a = \hat{V}(P_i)$ and $b = 0$ to (6.2) and (5.4) in this case.

We estimated Taylor's power law in the proportion of pecky rice grains that were primary caused by the rice leaf bug, *T. caelestialium*. The parameter b of Taylor's power law may change depending on the species; the parameter b is influenced by the dispersal rate and reproduction rate of each species (Yamamura 2000). The parameter b generally becomes close to 1 if the insect has a high tendency of dispersal, while it becomes close to 2 if the insect has a low tendency of dispersal (Yamamura 2000). Yamamura (2001) estimated the parameter b of the distribution of 4 insect pests that have quite different abilities of dispersal in cabbage fields. The estimated b for the green peach aphid *Myzus persicae*, eggs of the diamondback moth *Plutella xylostella*, eggs of the small white butterfly *Pieris rapae crucivora*, and eggs of the beet semi-looper *Autographa nigrisigna*, were 1.74, 1.52, 1.28, and 1.15, respectively. The decreasing order of the estimated size of b coincides with the increasing order of dispersal ability. Pecky rice grains in Japan are mostly caused by either of the three species of bugs: the rice leaf bug *T. caelestialium*, the sorghum plant bug *Stenotus rubrovittatus*, and the rice bug *Leptocorisa chinensis* (Watanabe and Higuchi 2006). The optimal sample size calculated in this paper will be applicable for most of the northern part of Japan (including Hokuriku, Tohoku, and Hokkaido districts) where *T. caelestialium* is the primary species. However, if we apply the method to the districts where *S. rubrovittatus* or *L. chinensis* are the primary species, it will be preferable to re-estimate the parameters of Taylor's power law for these species.

AKNOWLEDGMENTS

We would like to thank anonymous reviewers for their valuable comments that greatly improved the manuscript.

REFERENCES

Bar-Lev, S. K., Stadje, W., and Van der Duyn Schouten, F. A. (2006), "Group Testing Procedures with Incomplete Identification and Unreliable Testing Results," *Applied Stochastic Models in Business and Industry*, 22, 281--296.

Bhattacharyya, G. K., Karandinos, M. G., and DeFoliart, G. R. (1979), "Point Estimates and Confidence Intervals for Infection Rates Using Pooled Organisms in Epidemiologic Studies," *American Journal of Epidemiology*, 109, 124--131.

Boswell, M. T., Burnham, K. P., and Patil., G. P. (1988), "Role and Use of Composite Sampling and Capture-Recapture Sampling in Ecological Studies," in *Handbook of Statistics*, eds. P. R. Krishnaiah and C. R. Rao, North Holland: Elsevier Science, pp. 469--488.

Brookmeyer, R. (1999), "Analysis of Multistage Pooling Studies of Biological Specimens for Estimating Disease Incidence and Prevalence " *Biometrics*, 55, 608--612.

Brown, G. H., and Fisher, N. I. (1972), "Subsampling a Mixture of Sampled Material," *Technometrics*, 14, 663--668.

Burrows, P. M. (1987), "Improved Estimation of Pathogen Transmission Rates by Group Testing," *Phytopathology*, 77, 363--365.

Chen, C. L., and Swallow, W. H. (1990), "Using Group Testing to Estimate a Proportion, and to Test the Binomial Model," *Biometrics*, 46, 1035--1046.

Chiang, C. L., and Reeves, W. C. (1962), "Statistical Estimation of Virus Infection Rates in Mosquito Vector Populations," *American Journal of Hygiene*, 75, 377--391.

Chick, S. E. (1996), "Bayesian Models for Limiting Dilution Assay and Group Test Data," *Biometrics*, 52, 1055--1062.

Cochran, W. G. (1977), *Sampling Techniques* (3rd ed.), New York: Wiley.

Colón, S., Patil, G. P., and Taillie, C. (2001), "Estimating Prevalence Using Composites," *Environmental and Ecological Statistics*, 8, 213--236.

Dorfman, R. (1943), "The Detection of Defective Members of Large Populations," *Annals of Mathematical Statistics*, 14, 436--440.

Elder, R. S., Thompson, W. O., and Myers, R. H. (1980), "Properties of Composite Sampling Procedures," *Technometrics*, 22, 179--186.

Emmanuel, J. C., Bassett, T., M., Smith, H. J., and Jacob, J. A. (1988), "Pooling of Sera for HIV Testing: An Economical Method for Use in Developing Countries," *Journal of Clinical Pathology*, 41, 582--585.

Federal Grain Inspection Service (1995), *Grain Inspection Handbook, Book 1, Grain Sampling*, Washington, D.C.: United States Department of Agriculture.

Gastwirth, J. L., and Johnson, W. O. (1994), "Screening with Cost-Effective Quality Control: Potential Applications to HIV and Drug Testing," *Journal of American Statistical Association*, 89, 972--981.

Hsu, L. (2005), "Group Testing with a Goal in Estimating the Number of Defects under Imperfect Environmental Stress Screen Levels," *Communications in Statistics - Theory and Methods*, 34, 1363--1377.

Hughes-Oliver, J. M., and Swallow, W. H. (1994), "A Two-Stage Adaptive Group-Testing Procedure for Estimating Small Proportions," *Journal of American Statistical Association*, 89, 982--993.

ISO (1990), *ISO 542, Oilseeds -- Sampling*, Genève: International Organization for Standardization.

--- (1999), *ISO 13690, Cereals, Pulses and Milled Products -- Sampling of Static Batches*, Genève: International Organization for Standardization.

--- (2000), *ISO 10725, Acceptance Sampling Plans and Procedures for the Inspection of Bulk Material* Genève: International Organization for Standardization.

--- (2002), *ISO 6644, Flowing Cereals and Milled Cereal Products -- Automatic Sampling by Mechanical Means*, Genève: International Organization for

Standardization.

--- (2003), *ISO 11648-1, Statistical Aspects of Sampling from Bulk Materials -- Part 1: General Principles*, Genève: International Organization for Standardization.

Iwasaki, M. (2005), "Rule of 3 and Related Topics," *Proceedings of the 2005 Symposium of the Biometric Society of Japan*, 1--2 (in Japanese).

Japan Plant Protection Association (2003), *Handbook of Experiments for Practical Use of Pesticides*, Tokyo: Japan Plant Protection Association (in Japanese).

Japanese Ministry of Health Labor and Welfare (2001), "Inspection Method of Food Using Genetically Modified Organisms, Notification No. 110, Annex," (in Japanese).

Johnson, G. D., and Patil, G. P. (2001), "Cost Analysis of Composite Sampling for Classification," *Environmental and Ecological Statistics*, 8, 91--107.

Johnson, N. L., Kotz, S., and Kemp, A. W. (2005), *Univariate Discrete Distrbutions* (3rd ed.), New York: Wiley.

Jovanovic, B. D., and Levy, P. S. (1997), "A Look at the Rule of Three," *American Statistician*, 51, 137--139.

Kiritani, K. (2006), "Predicting Impacts of Global Warming on Population Dynamics and Distribution of Arthropods in Japan " *Population Ecology*, 48, 5--12.

Kuno, E. (1986), *Research Methods for Population Dynamics of Animals: Estimation Methods of Populations*, Tokyo: Kyoritsu (in Japanese).

--- (1991), "Verifying Zero-Infestation in Pest Control: A Simple Sequential Test Based on the Succession of Zero Sample," *Researches on Population Ecology*, 33, 29--32.

Lancaster, V., and Keller-McNulty, S. (1998), "A Review of Composite Sampling Methods," *Journal of American Statistical Association*, 93, 1216--1230.

Lohr, S. L. (1999), *Sampling: Design and Analysis*, Pacific Grove: Duxbury Pr.

Lovinson, G., Gore, S. D., and Patil, G. P. (1994), "Design and Analysis of Composite

Sampling Procedures: A Review," in *Handbook of Statistics*, eds. G. P. Patil and C. R. Rao, Amsterdam: Elsevier, pp. 103--166.

Ministry of Agriculture Forestry and Fisheries (2001), "Regulations for the Agricultural Product Standards (Notification No. 244 of the Ministry of Agriculture, Forestry and Fisheries of Japan)," http://www.kokuji.maff.go.jp/kokujituti/top.asp (in Japanese).

Minotani, C. (2003), *Handbook of Statistical Distributions*, Tokyo: Asakura (in Japanese).

Patil, G. P. (2002), "Composite Sampling," in *Encyclopedia of Environmetrics*, eds. A. H. El-Shaarawi and W. W. Piegorsch, Chichester: Wiley, pp. 387--391.

Perry, J. N. (1981), "Taylor's Power Law for Dependence of Variance on Mean in Animal Populations," *Journal of the Royal Statistical Society Series C - Applied Statistics*, 30, 254--263.

Rohde, C. (1976), "Composite Sampling," *Biometrics*, 32, 273--282.

Rohlf, F. J., Akçakaya, H. R., and Ferraro., S. P. (1996), "Optimizing Composite Sampling Protocols," *Environmental Science and Technology*, 30, 2899--2905.

Shimizu, K. (2006), *Insurance Risk Models: Calculating Claims Distributions*, Tokyo: Kyoritsu (in Japanese).

Sterrett, A. (1957), "On the Detection of Defective Members of Large Populations," *Annals of Mathematical Statistics*, 28, 1033--1036.

Swallow, W. H. (1985), "Group Testing for Estimating Infection Rates and Probabilities of Disease Transmission," *Phytopathology*, 75, 882--889.

--- (1987), "Relative Mean Squared Error and Cost Considerations in Choosing Group Size for Group Testing to Estimate Infection Rates and Probabilities of Disease Transmission," *Phytopathology*, 77, 1376--1381.

Taylor, L. R. (1961), "Aggregation, Variance and the Mean," *Nature*, 189, 732--735.

Taylor, L. R., Woiwod, I. P., and Perry, J. N. (1978), "The Density-Dependence of

Spatial Behaviour and the Rarity of Randomness," *Journal of Animal Ecology*, 47, 383--406.

--- (1979), "The Negative Binomial as a Dynamic Ecological Model for Aggregation, and the Density Dependence of *k*," *Journal of Animal Ecology*, 48, 289--304.

Taylor, L. R. (1984), "Assessing and Interpreting the Spatial Distribution of Insect Populations," *Annual Review of Entomology*, 29, 321--357.

Tebbs, J. M., Bilder, C. R., and Moser, B. K. (2003), "An Empirical Bayes Group-Testing Approach to Estimating Small Proportions," *Communications in Statistics - Theory and Methods*, 32, 983--995.

Thompson, K. H. (1962), "Estimation of the Proportion of Vectors in a Natural Population of Insects," *Biometrics*, 18, 568--578.

Thompson, S. K. (2002), *Sampling* (2nd ed.), New York: Wiley.

United States Environmental Protection Agency (2000), *Guidance for Choosing a Sampling Design for Environmental Data Collection (EPA QA/G-5S)*: United States Environmental Protection Agency.

van Belle, G. (2002), *Statistical Rules of Thumb*, New York: Wiley.

Watanabe, T., and Higuchi, H. (2006), "Recent Occurrence and Problem of Rice Bugs," *Shokubutsu Boeki (Plant Protection)*, 60, 201--203 (in Japanese).

Yamamura, K., and Sugimoto, T. (1995), "Estimation of the Pest Prevention Ability of the Import Plant Quarantine in Japan," *Biometrics*, 51, 482--490.

Yamamura, K. (2000), "Colony Expansion Model for Describing the Spatial Distribution of Populations," *Population Ecology*, 42, 161--169.

--- (2001), "Theoretical Studies on the Relationship between the Spatial Distribution of Insects and Population Dynamics," *Bulletin of the National Institute for Agro-Environmental Sciences*, 19, 1--60 (in Japanese with English summary).

Yamamura, K., Yokozawa, M., Nishimori, M., Ueda, Y., and Yokosuka, T. (2006), "How to Analyze Long-Term Insect Population Dynamics under Climate Change:

50-Year Data of Three Insect Pests in Paddy Fields," *Population Ecology*, 48, 31--48.

Yamamura, K., and Hino, A. (2007), "Estimation of the Proportion of Defective Units by Using Group Testing under the Existence of a Threshold of Detection," *Communications in Statistics - Simulation and Computation*, 36, 949--957.

Zenios, S. A., and Wein, L. M. (1998), "Pooled Testing for HIV Prevalence Estimation: Exploiting the Dilution Effect " *Statistics in Medicine*, 17, 1447--1467.

FIGURE CAPTIONS

Figure 1. Taylor's power law between the mean and variance of the proportion of pecky rice grains. Three rice varieties are indicated by different symbols. The linear regression was calculated including three rice varieties: $\log_e[\hat{V}(P_i)] = -2.19 + 1.60 \log_e(\hat{P}_0)$, $R^2 = 0.91$.

Figure 2. Sample size to achieve a given relative precision (D). The curves indicate the combination of n_1 and n_2 that achieve $D < 0.25$ for all P_0 in the range of $P_0 > P_c$. Five curves for different values of P_c are shown. The solid circle indicates the optimal combination of n_1 and n_2 that was calculated by using (6.2) for $P_c = 0.001$ and $(c_1/c_2) = 500$. The broken line indicates a slope of −500. The shaded area indicates the nonexistent combination of n_1 and n_2 where the required number of drawn grains exceeds the total number of drawn grains, i.e., the region of $n_2 > sn_1$. The following parameters were used: $s = 1400$, $a = \exp(-2.19)$, and $b = 1.60$.

Figure 3. Sample size to achieve a given consumer's risk (β). The curves indicate the combination of n_1 and n_2 that achieve $\beta < 0.05$ for all P_0 in the range of $P_0 > P_c$. Five curves for different values of P_c are shown. The solid circle indicates the optimal combination of n_1 and n_2 that was calculated by using (8.4) for $P_c = 0.001$ and $(c_1/c_2) = 500$. The broken line indicates a slope of −500. The meaning of shaded area is the same as that in Figure 2. The following parameters were used: $s = 1400$, $a = \exp(-2.19)$, and $b = 1.60$.

Figure 4. Conceptual difference between two kinds of control, the control of precision and control of consumer's risk. The operating characteristic curves (OC-curves)

are compared under the assumption that the probability that a rice grain around the ith plant is pecky (P_i) fluctuates following a gamma distribution spatially. The broken curve indicates the OC-curve for the zero-tolerance method using P_c = 0.001 and $\beta = 0.05$, i.e., $n_1^{\#} = 5.5$ and $n_2^{\#} = 5001$. The solid curve indicates the OC-curve for the rejection procedure where we reject the product if we obtain $\hat{P}_0 > P_c$ under the sampling procedure that controls $D = 0.25$ for $P_c = 0.001$, i.e, $n_1^* = 58$ and $n_2^* = 31000$. The following parameters were used: $a = \exp(-2.19)$ and $b = 1.60$.

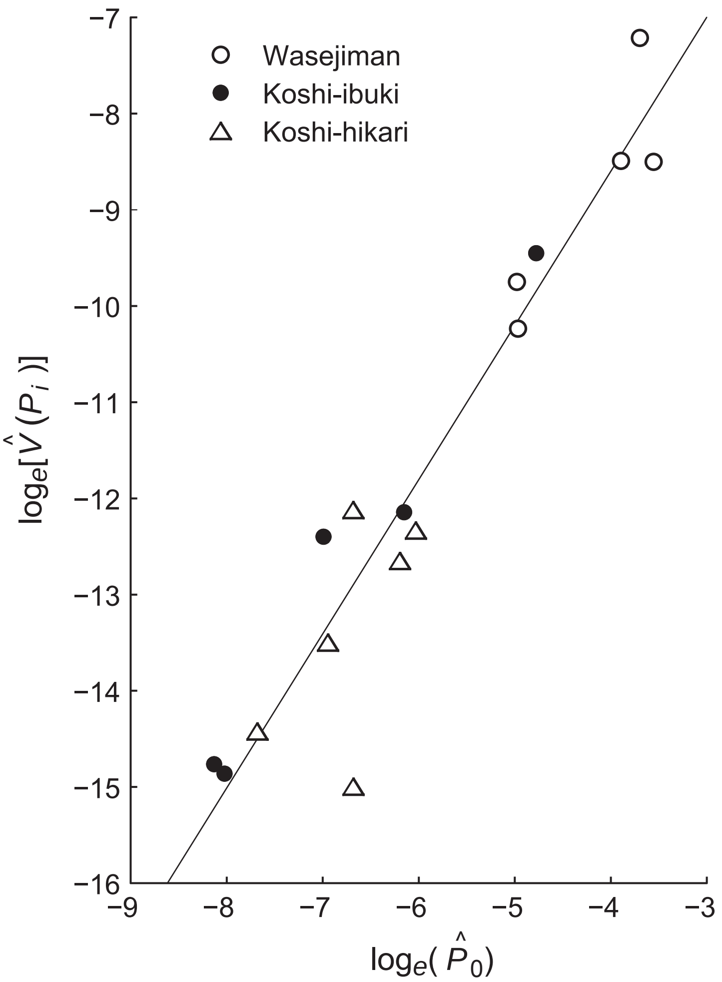
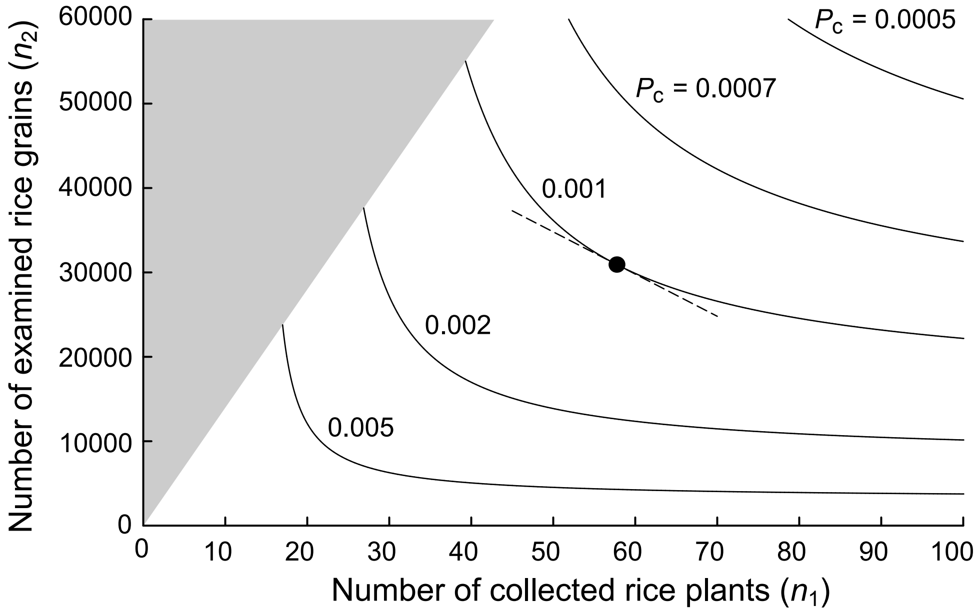
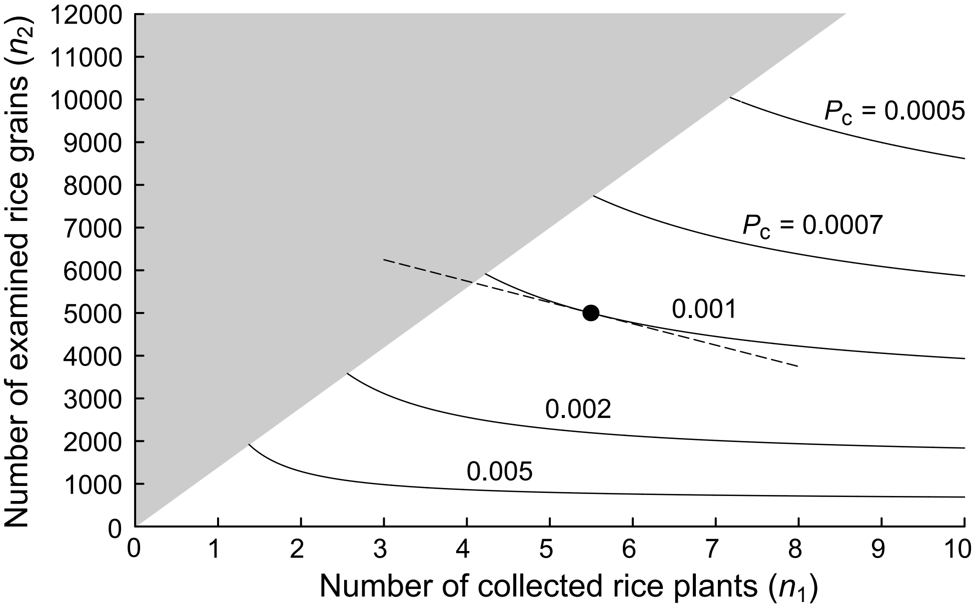

Wasejiman
Koshi-ibuki
Koshi-hikari
−7
−8
−9
−10
−11
−12
−13
−14
−15
−16
−9
−8
−7
−6
−5
−4
−3
$\log_e[\hat{V}(P_i)]$
$\log_e(\hat{P}_0)$

Figure 1



60000
50000
40000
30000
20000
10000
0
0
10
20
30
40
50
60
70
80
90
100
Number of examined rice grains (n_2)
Number of collected rice plants (n_1)
P_c = 0.0005
P_c = 0.0007
0.001
0.002
0.005

Figure 2



Number of examined rice grains (n_2)
Number of collected rice plants (n_1)
P_c = 0.0005
P_c = 0.0007
0.001
0.002
0.005

Figure 3

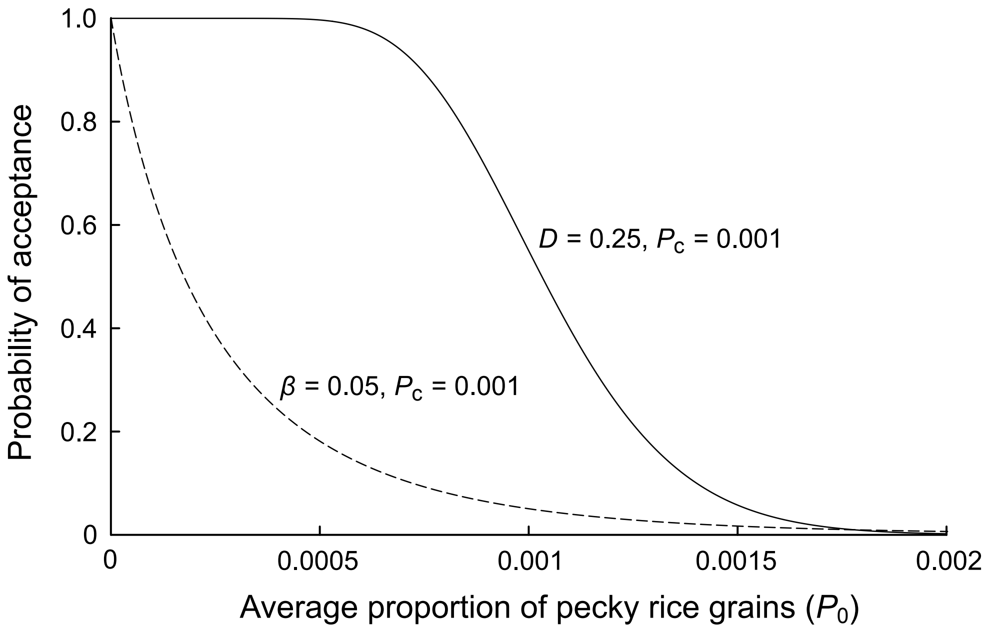

1.0
0.8
0.6
0.4
0.2
0
Probability of acceptance
D = 0.25, Pc = 0.001
β = 0.05, Pc = 0.001
0
0.0005
0.001
0.0015
0.002
Average proportion of pecky rice grains (P0)

Figure 4