ESTIMATION OF THE PROPORTION OF DEFECTIVE UNITS BY USING GROUP TESTING UNDER THE EXISTENCE OF A THRESHOLD OF DETECTION

Kohji Yamamura [1] and Akihiro Hino [2]

[1] 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

[2] GM Analytical Evaluation Research Team and Gustatory Biology Laboratory, National Food Research Institute, 2-1-12 Kannondai, Tsukuba 305-8642, Japan

Key Words: Confidence intervals; genetically modified organisms; group testing; risk management; threshold of detection.

Received 3 March 2006; Accepted 22 January 2007

ABSTRACT

Group testing procedures, in which groups containing several units are tested without testing each unit, are widely used as cost-effective procedures in estimating the proportion of defective units in a population. A problem arises when we apply these procedures to the detection of genetically modified organisms (GMOs), because the analytical instrument for detecting GMOs has a threshold of detection. If the group size (i.e., the number of units within a group) is large, the GMOs in a group are not detected due to the dilution even if the group contains one unit of GMOs. Thus, most people conventionally use a small group size (which we call conventional group size) so that they can surely detect the existence of defective units if at least one unit of GMOs is included in the group. However, we show that we can estimate the proportion of defective units for any group size even if a threshold of detection exists; the estimate of the proportion of defective units is easily obtained by using functions implemented in a spreadsheet. Then, we show that the conventional group size is not always optimal in controlling a consumer's risk, because such a group size requires a larger number of groups for testing.

1. INTRODUCTION

Estimation of the proportion of defective (unsatisfactory) units in a population is an important part of the process of risk assessment of agricultural products. In some cases, testing the units one-by-one is inefficient, especially when very few of them are defective and they are cheap relative to the cost of the test. In such cases, it is often preferable to form groups of units, and test all units in a group simultaneously. This procedure is usually called "group testing" in the statistical literature (Chen and Swallow, 1990). An early application of group testing was to estimate the prevalence of plant virus transmission by insects (Watson, 1936; Thompson, 1962). This procedure was later applied in various studies in the fields including phytopathology, public health, and plant quarantine (Chiang and Reeves, 1962; Bhattacharyya et al., 1979; Swallow, 1985; Romanow et al., 1986; Burrows, 1987; Swallow, 1987; Yamamura and Sugimoto, 1995; Hughes and Gottwald, 1998; Zenios and Wein, 1998; Remund et al., 2001; Xie et al., 2001).

The unintentional mingling of genetically modified organisms (GMOs) with non-GMOs has recently occupied considerable public attention. We have to urgently construct the official procedure of sampling inspection to meet the requirements of legislation. Group testing is useful for this inspection, because the proportion of mingling is usually very small. However, we meet difficulties in applying group testing. We must first determine two parameters in applying group testing procedures: the number of units within a group, which is denoted by n, and the number of groups in the testing procedure, which is denoted by w. We are using ELISA or PCR (polymerase chain reaction) in detecting GMOs. These analytical instruments, as well as other instruments, have their threshold of detection below which the existence of the subject material is not detected. If the group size (n) is large, the material will not be detected even if the group contains one defective unit, because the material is highly diluted. The problem of the threshold of detection has not yet been fully considered, although several authors discussed the optimal allocation of n and w (Thompson, 1962; Swallow, 1985; Swallow, 1987). Most people conventionally use a small group size (which we call conventional group size) so that they can surely detect the existence of defective units

if at least one unit of GMOs is included in the group. In this paper, we show that we can estimate the proportion of defective units for any group size even if a threshold of detection exists; the estimate of the proportion of defective units is easily obtained by using functions implemented in a spreadsheet. Then, we show that the conventional group size is not always optimal in controlling a consumer's risk, because such a group size requires a larger number of groups for testing. We apply this procedure to the detection of genetically modified corn. We provide an example of Excel spreadsheet to calculate the following two quantities: (1) the maximum likelihood estimate of the proportion of defective units, and (2) the sample size to satisfy a given consumer's risk.

2. MAXIMUM LIKELIHOOD ESTIMATE AND CONFIDENCE INTERVALS

Let p be the true proportion of defective units, n be the number of units in a group, w be the total number of groups to be tested, and v be the number of groups that were judged as defective. We assume that the sampling units are drawn at random and mixed uniformly within each group. If there is no threshold of detection, the likelihood is given by

$$L(p|\, n, w, v) = \left(\begin{array}{c} w \\ v \end{array} \right) \{(1-p)^n\}^{(w-v)} \{1-(1-p)^n\}^v. \tag{2.1}$$

The maximum likelihood estimate of p that satisfy $\partial \log_e(L)/\partial p = 0$ is given by

$$\hat{p} = 1 - \left(1 - \frac{v}{w}\right)^{\frac{1}{n}}. \tag{2.2}$$

This is the conventional estimator for group testing with groups of equal size. Now we assume the existence of threshold detection. Let q be the threshold proportion of detection, and k be the threshold number of defective units that enables the detection. If $q > 0$, the quantity of k is given by $\lceil nq \rceil$, that is, the smallest integer not smaller than nq. If $q = 0$, k equals 1. The probability that a group contains i defective units is given by a binomial distribution,

$$\left(\begin{array}{c} n \\ i \end{array} \right) p^i (1-p)^{n-i}. \tag{2.3}$$

Let $F(x|\, \alpha, \beta)$ be the distribution function of the beta distribution with parameters α and β. Let $F^{-1}(x|\, \alpha, \beta)$ be the inverse function. The probability that a group is judged as

defective is given by the probability that a group contains k or more defective units. We can derive the probability by using the relation between the binomial distribution and the beta distribution,

$$\sum_{i=k}^{n} \left(\begin{array}{c} n \\ i \end{array} \right) p^i (1-p)^{n-i} = F(p|\, k, n-k+1). \tag{2.4}$$

Then, the likelihood is given by

$$L(p|\, n, k, w, v) = \left(\begin{array}{c} w \\ v \end{array} \right) \{1 - F(p|\, k, n-k+1)\}^{(w-v)} \{F(p|\, k, n-k+1)\}^v . \tag{2.5}$$

This quantity is maximized at $F(p|\, k, n-k+1) = v/w$. Hence, the maximum likelihood estimate is given by

$$\hat{p} = F^{-1}((v/w)|\, k, n-k+1). \tag{2.6}$$

This quantity simplifies to Eq. (2.2) when $k = 1$. We can calculate $\hat{p}$ by using the BETAINV function or the FINV function of Excel with the statements IF(v=0, 0, IF(v/w=1, 1, BETAINV(v/w, k, n-k+1))).

Considerable amount of debate has been conducted on the confidence intervals for a binomial parameters (Clopper and Pearson, 1934; Sterne, 1954; Newcombe, 1998; Brown et al., 2001; Henderson and Meyer, 2001; Reiczigel, 2003). As for most continuous variables, we can calculate the confidence intervals by using the inverse of testing (Mood et al., 1974). As for discrete variables such as binomial variables, however, the coverage probability of the confidence intervals constructed by the inverse of exact testing, which is conventionally called "exact confidence intervals", is usually larger than $100(1-\alpha)\%$. Hepworth (1996) calculated exact confidence intervals for group testing while Tebbs and Bilder (2004) and Hepworth (2004) compared several confidence interval methods. In our paper, however, we use exact confidence intervals because of the ease of calculation; we can calculate them by simple equations. The following equation gives the lower limit (p_L) of the two-sided $100(1-\alpha)\%$ exact confidence interval,

$$\sum_{i=v}^{w} L(p_L|\, n, k, w, i) = \frac{\alpha}{2}. \tag{2.7}$$

By using a relation similar to Eq. (2.4), we can rewrite the above equation as

$$F(F(p_L|\, k, n-k+1)|v, w-v+1) = \frac{\alpha}{2}. \tag{2.8}$$

Therefore,

$$p_L = F^{-1}(F^{-1}(\,(\alpha/2)|\, v, w-v+1)|k, n-k+1). \tag{2.9}$$

Similarly, the upper limit (p_U) of the two-sided $100(1-\alpha)\%$ exact confidence interval, is

$$p_U = F^{-1}(F^{-1}(1-(\alpha/2)|\, v+1, w-v)|k, n-k+1). \tag{2.10}$$

We can calculate p_L and p_U by using Excel with the statements IF(v>0, BETAINV(BETAINV(alpha/2, v, w-v+1), k, n-k+1), 0) and IF(v/w=1,1, BETAINV(BETAINV(1-alpha/2, v+1, w-v), k, n-k+1)), respectively, where k is given by IF(q>0, CEILING(n*q,1), 1). An example of Excel spreadsheet is given by <http:// cse.niaes.affrc.go.jp/yamamura/ Estimation_of_proportion.xls>.

3. REQUIRED SAMPLE SIZE

We next consider the inspection of consignments where the consignment is declared as unsatisfactory if at least one group is judged as defective, that is if $v > 0$. This type of judgment is commonly adopted in plant quarantine inspection because this judgment usually requires the smallest sample size; A judgment using a larger threshold, such as $v > 1$ and $v > 2$, requires a larger sample size. Let p_c be the threshold proportion of defective units above which consumers cannot tolerate. A consignment having a proportion of defectives p is defined as unsatisfactory if $p_c < p$. Let us consider a risk management procedure where an unsatisfactory consignment is accepted with a probability less than β. This probability is usually called "consumer's risk". Then, the consumer's requirement is expressed by the following inequality by substituting $v = 0$ into Eq. (2.5),

$$\{1 - F(p|\, k, n-k+1)\}^w < \beta. \tag{3.1}$$

Thus, we have

$$w > \frac{\log_e(\beta)}{\log_e\{1 - F(p|\, k, n-k+1)\}}. \tag{3.2}$$

This inequality must be satisfied for all p within the range $p_c < p \leq 1$. Let w_c be the minimum sample size that satisfies the above inequality for all p within the range $p_c < p \leq 1$. The quantity of $F(p|\, k, n-k+1)$, that is the probability that a group contains k or more defective units (Eq. 2.4), decreases with decreasing p. Thus, the right hand side of Eq. (3.2) increases with decreasing p. Therefore, we replace '>' in Eq. (3.2) by '$\geq$' and replace p by p_c. Then, w_c is given by using a ceiling function that indicates the minimum integer that satisfy this modified version of inequality.

$$w_c = \left\lceil \frac{\log_e(\beta)}{\log_e\{1 - F(p_c|\, k, n-k+1)\}} \right\rceil . \tag{3.3}$$

We can calculate the required sample size by using the BETADIST function of Excel with the expression CEILING(LOG(beta)/LOG(1-BETADIST(pc,k,n-k+1)),1), where k is given by IF(q>0,CEILING(n*q,1),1). An example of Excel spreadsheet is given by <http:// cse.niaes.affrc.go.jp/yamamura/Estimation_of_proportion.xls>. If $k = 1$, we obtain a simple formula as an approximation,

$$nw_c = \log_e(\beta)/\log_e(1 - p_c). \tag{3.4}$$

This formula becomes another well-known formula, if p_c is sufficiently small,

$$nw_c = -\log_e(\beta)/p_c. \tag{3.5}$$

4. APPLICATION

We apply the above procedure to the detection of genetically modified corn CBH351 (StarLink^{TM}; a variety of Bt corn). The unit of sampling is one grain of corn in this case. We can use both ELISA and PCR for the detection of CBH351 in the protocol prescribed by the Japanese Ministry of Agriculture, Forestry and Fisheries. When we use the primer pairs designed by Matsuoka et al. (2001), the threshold of detection is $q = 0.0005$; the threshold have been examined by performing 'blind tests' among several different laboratories. The hypothetical example in Table 1 indicates several characteristics of estimates. In Case1, Case2, and Case3, we increased the number of units per group (n) while keeping w at 10

and v at 1. The maximum likelihood estimate increased with increasing n in this series of example. In contrast, in Case4, Case5, and Case6 where we increased n while setting $w = 10$ and $v = 5$, the maximum likelihood estimate decreased with increasing n. The width of the 95% two-sided confidence interval decreased with increasing n in both series of examples.

US Federal Grain Inspection Service has adopted 2400 grains as the standard sample size for CBH351 testing (USDA Federal Grain Inspection Service, 2001). Equation (3.5) indicates that this sample size corresponds to the case of $\beta = 0.05$ and $p_c = 0.00125$. Then, for this combination of β and p_c, we calculated the change in the required number of groups (w_c) that occurs with the increase in group size (n). The upper panel of Fig. 1 indicates that the required number of groups decreases with increasing n by a complicated manner when $q = 0.0005$; it jumps at the position where the threshold number of defective units (k) changes. According to the law of large numbers, the proportion of defective units in a group converges to p as the group size increases. If the threshold of detection (q) is smaller than p_c, the defectiveness of a group is detected by a probability approaching 1 in this case. Hence, the required number of groups approaches 1 with increasing n if $q < p_c$. On the other hand, if $q > p_c$, the required number of groups generally increases as exemplified by the lower panel of Fig. 1.

5. DISCUSSION

Most people conventionally use a smaller group size (n) so that they can more surely detect the existence of defective units in the group. The upper panel of Fig. 1 indicates that such a testing procedure is not always optimal if $q < p_c$. We must examine 2400 units if $\beta = 0.05$ and $p_c = 0.00125$ as stated before. When the threshold of detection (q) is 0.0005, we cannot detect the defectiveness of a group even if the group contains a defective unit within the 2400 units if we test all units as one group. Hence, we usually divide the 2400 units into half, that is, we adopt $n = 1200$ and $w_c = 2$. However, if we use the hollow near the arrow (B) in the upper panel of Fig. 1 ($n = 4000$, for example), the required number of groups is 1. Thus, we can control the consumer's risk by a single testing procedure. The reduction in the required number of group (w_c) will be important if the cost of testing a

group is much larger than that of preparing a laboratory sample.

We have assumed that the defectiveness is never detected under the threshold of detection. However, there may be some intermediate concentration where the defectiveness is detected with a small probability. Hung and Swallow (1999) considered a model called DE1 where the probability of detection continuously decreases with decreasing concentration under a threshold. They numerically determined that we can reduce the mean squared error (MSE) by using group testing even under the existence of a threshold when p_c is small. In our model, if the probability of detection is not zero under the threshold of detection q, the probability of detection may become larger than the quantity given by Eq. (2.4). Hence, the sample size required for the detection (w_c) may become smaller than the quantity given by Eq. (3.3). Thus, the consumer's requirement will be still satisfied if we use the sample size given by Eq. (3.3) although there may be some loss of efficiency.

In several procedure of sampling, samples are drawn by clusters, which are called "increments", from each consignment; that is, more than one unit is drawn from a point of a consignment. Such a sampling procedure causes difficult problems, because most equations for group testing are derived from the assumption of random sampling. If the defective units are randomly located in each consignment, the equation derived from the assumption of random sampling is still applicable. If the spatial distribution of defective units is aggregated, however, the assumption of random sampling will yield several problems in the variance of estimates. We must estimate the variance between increments to estimate exact variance in this case. This will be one of the problems that should be solved in the future studies.

REFERENCES

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. *Amer. J. Epidemiol.*, **109**, 124–131.

Brown, L. D., Cai, T. T., and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.*, **16**,101–133.

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. *Amer. J. Hygiene*, **75**, 377–391.

Clopper, C. J., and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, **26**, 404–413.

Henderson, M., and Meyer, M. C. (2001). Exploring the confidence interval for a binomial parameter in a first course in statistical computing. *Amer. Stat.*, **55**, 337–344.

Hepworth, G. (1996). Exact confidence intervals for proportions estimated by group testing. *Biometrics*, **52**, 1134–1146.

Hepworth, G. (2004). Mid-P confidence intervals based on the likelihood ratio for proportions estimated by group testing. *Aust. and New Zealand J Statist.*, **46**, 391–405.

Hughes, G., and Gottwald, T. R. (1998). Survey methods for assessment of citrus tristeza virus incidence *Phytopathology*, **88**, 715–723.

Hung, M., and Swallow, W. H. (1999). Robustness of group testing in the estimation of proportions. *Biometrics*, **55**, 231–237.

Matsuoka, T., Kuribara, H., Suefuji, S., Miura, H., Kusakabe, Y., Akiyama, H., Goda, Y., Isshiki, K., Toyoda, M., and Hino, A. (2001). A detection method for recombinant DNA from genetically modified maize CBH351. *Shokuhin Eiseigaku Zasshi (J Food Hyg. Soc. Japan)*, **42**, 197–201.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the theory of statistics*, New York: McGraw-Hill.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat. Med.*, **17**, 857–872.

Reiczigel, J. (2003). Confidence intervals for the binomial parameter: some new considerations. *Stat. Med.*, **22**, 611–621.

Remund, K. M., Dixon, D. A., Wright, D. L., and Holden, L. R. (2001). Statistical considerations in seed purity testing for transgenic traits. *Seed Sci. Res.*, **11**, 101–120.

Romanow, L. R., Moyer, J. W., and Kennedy, G. G. (1986). Alteration of efficiencies of acquisition and inoculation of watermelon mosaic virus 2 by plant resistance to the virus and to an aphid vector. *Phytopathology*, **76**, 1276–1281.

Sterne, T. (1954). Some remarks on confidence or fiducial limits. *Biometrika*, **41**, 275–278.

Swallow, W. H. (1985). Group testing for estimating infection rates and probabilities of disease transmission. *Phytopathology*, **75**, 882–889.

Swallow, W. H. (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.

Tebbs, J. M., and Bilder, C. R. (2004). Confidence interval procedures for the probability of disease transmission in multiple-vector-transfer designs. *J. Agric. Biol. Environm. Statist.*, **9**, 75–90.

Thompson, K. H. (1962). Estimation of the proportion of vectors in a natural population of insects. *Biometrics*, **18**, 568–578.

USDA Federal Grain Inspection Service (2001). GIPSA Directive 9181.1, Testing for StarLink-Corn - lateral flow test strip method, <http://151.121.3.117/reference-library/directives/9181-1.pdf>.

Watson, M. A. (1936). Factors affecting the amount of infection obtained by aphis transmission of the virus Hy. III. *Philos. Trans. R. Soc. Lond., Ser. B*, **226**, 457–489.

Xie, M., Tatsuoka, K., Sacks, J., and Young, S. (2001). Group testing with blockers and synergism. *J. Amer. Stat. Assoc.*, **96**, 92–102.

Yamamura, K., and Sugimoto, T. (1995). Estimation of the pest prevention ability of the import plant quarantine in Japan. *Biometrics*, **51**, 482–490.

Zenios, S. A., and Wein, L. M. (1998). Pooled testing for HIV prevalence estimation: exploiting the dilution effect. *Stat. Med.*, **17**, 1447–1467.

TABLES

Table 1. Hypothetical example of the estimation of the proportion of transgenic corn gene CBH351.

	Case1	Case2	Case3	Case4	Case5	Case6
Assumption						
Number of grains in a group (n)	1000	3000	10000	1000	3000	10000
Threshold proportion of detection (q)	0.05%	0.05%	0.05%	0.05%	0.05%	0.05%
Minimum number of defective grains for the detection (k)	1	2	5	1	2	5
Total number of tested groups (w)	10	10	10	10	10	10
Number of defective groups (v)	1	1	1	5	5	5
Estimates of the proportion of defective grains						
Maximum likelihood estimate	0.0105%	0.0177%	0.0243%	0.0693%	0.0559%	0.0467%
Two-sided 95% upper confidence limit	0.0589%	0.0503%	0.0438%	0.1675%	0.1027%	0.0685%
Two-sided 95% lower confidence limit	0.0003%	0.0024%	0.0092%	0.0207%	0.0263%	0.0301%
One-sided 95% upper confidence limit	0.0501%	0.0453%	0.0412%	0.1502%	0.0950%	0.0651%

FIGURE LEGENDS

Figure 1. Irregular changes in the number of required groups (w_c) that occur with increasing group size (n) ($\beta = 0.05$ and $p_c = 0.00125$). Upper panel: the threshold of detection is set at $q = 0.0005$. Arrows indicate the position where the threshold number of defective units (k) changes. A: $k = 1$ to 2, B: $k = 2$ to 3, C: $k = 3$ to 4. Lower panel: the threshold of detection is set at $q = 0.0025$.

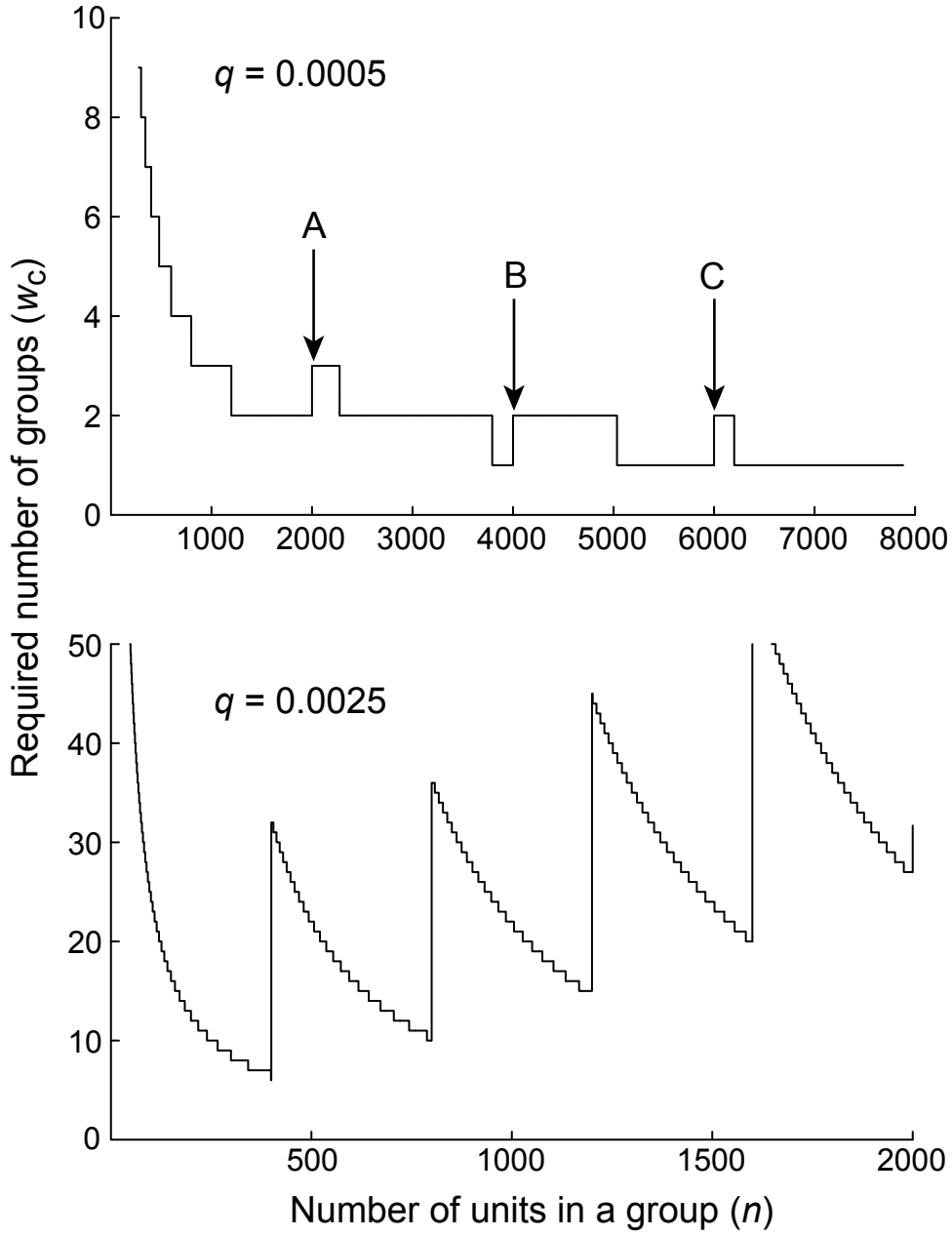
Required number of groups (w_c)
q = 0.0005
A
B
C
q = 0.0025
Number of units in a group (n)

Figure 1