

SYSTEMATIC VERSUS SIMPLE RANDOM SAMPLING IN PLANT PROTECTION

S.P. WORNER¹, R. SHAH² and R.B. CHAPMAN¹

¹*Ecology and Entomology Group, Lincoln University, P.O. Box 84, Canterbury*

²*ARI - Tarnab, Peshawar, Pakistan*

ABSTRACT

Systematic sampling methods are often used in disease and pest control studies because they are easy to implement and give uniform coverage of sample units over the population to be sampled. While there are few criticisms of these methods, computer simulations show that systematic sampling should be avoided when sampling any characteristic that is patchily dispersed throughout a population of sample units. Unless large sample sizes can be used when patchiness or contagion is suspected, simple random sampling is recommended. A QBASIC computer program that can reduce the time taken to complete a simple random sample is briefly described.

Keywords: sampling, simulation, simple random, systematic, patchy dispersion.

INTRODUCTION

The effect of the dispersion pattern of a population characteristic to be sampled on the precision and accuracy of a sampling method used, is rarely considered. Many factors may influence the dispersion pattern of individuals in space or individuals with certain characteristics within a population. Migration patterns of insects and the history of insecticide use may cause insects and insect phenotypes to be over-dispersed in certain areas (eg. Brun and Suckling 1992). Diseases in plants may develop and spread from individual foci within a crop (eg. Basu *et al.* 1977). Parasitoids may concentrate activities near supplementary nectar and pollen resources near crop margins (eg. Landis and Haas 1992). To avoid bias, a simple random sampling method or a systematic sample that starts from an initial random point, is suggested (Manly 1992). While systematic schemes within different disciplines in plant protection may vary (random starting points may or may not be used), all are considered advantageous. Compared with simple random samples, systematic samples are easy to implement, take less time to complete and give a more even distribution of sample units throughout the population of interest. The only major caution concerning systematic sampling is the possibility that periodic variation in some environmental variable coincides with each sample unit, thereby biasing population estimates (Krebs 1989; Manly 1992). While many studies have compared the two sampling approaches in the field, computer simulation allows a more direct comparison of the efficiency of the two sampling methods using a population(s) with precisely known characteristics.

MATERIALS AND METHODS

A computer program written in Microsoft FORTRAN was used to simulate a sampling frame of 100 x 1000 cells (100,000 individuals) in a matrix. This rectangular sampling frame was specifically chosen so that hard copies could be examined to detect programming errors. In the matrix, individuals that possess a particular trait (eg. those that are resistant, diseased or parasitised, hereafter referred to as "positive individuals") were assigned the value 1. Individuals that did not possess the trait were represented by 0. Positive individuals were allocated either randomly throughout the matrix (using a uniform random number function to select the coordinates of each cell) or in clumps or patches. Where a patchy dispersion was used, each patch was

positioned randomly in the data matrix and the number of positive individuals within each patch was determined using a log-normal distribution function. The percentage frequency of positive individuals allocated to the matrix or grid was either 10 or 20% (10000 and 20000 individuals respectively). For each of these frequencies different dispersion patterns were used. These included, no patch (randomly dispersed individuals), 50, 30, 20, 10, 5 and 2 (highly contagious dispersion) patches of positive individuals. For every combination of the frequency of positive individuals, dispersion pattern and sampling plan, the sampling simulation was repeated 100 times using the same matrix or grid. Sample sizes from 50 to 1500 were compared for each combination.

For random sampling, the program selected the coordinates of each cell sampled using a random number function. For systematic sampling, cells were selected systematically starting from a single randomly selected cell. Thereafter, every n 'th cell was selected as the next sample unit. The size of n was based on the sample size and calculated in such a way to distribute the sample points in both x and y directions throughout the matrix. The numbers of positive individuals and the zero individuals were recorded by the program. To test the efficiency of the two sampling methods, the probability of detecting 90% of the actual number of positive individuals present in the population was determined. For example, for a sample size of 100, to detect 90% of the 10% positive individuals in the population, at least nine positive individuals in the sample are expected. In addition, the minimum and maximum numbers detected, standard errors, 95% confidence intervals and Lloyd's index of patchiness (Pedigo and Buntin 1994), I were calculated. The relationship between I and probability of detection was tested using Spearman's rank correlation (Battacharyya and Johnson 1977). Theoretical probabilities of detecting 90% of the positive individuals in the population were calculated using the cumulative binomial probability distribution function in Minitab version 9 (Minitab Inc.).

RESULTS

The effect of an increase in the degree of patchiness on the probability of detecting 90% of the positive individuals present in the population, using a sample size of 50, is shown in Table 1. Using random sampling for both the 10 and 20% frequencies, the efficiency of detection was not affected by changes in the degree of patchiness (as indicated by increasing values of I). When systematic sampling was used, there was a significant decrease ($P < 0.001$ for both frequencies) in the probability of detection of 90% of the positive individuals as patchiness increased. The probability of detecting 90% of the positive individuals at a frequency of 10 and 20% dispersed in two patches was only 0.18 and 0.23, respectively.

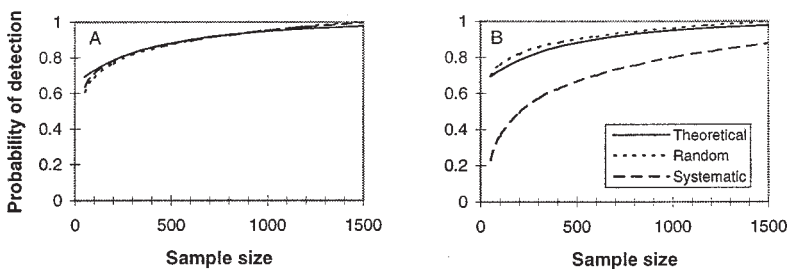


FIGURE 1: Probability of detecting 90% of positive individuals present in a population at a frequency of 20% for, (A) randomly dispersed individuals and, (B) where the individuals are dispersed in 5 patches. Theoretical probabilities of detection were calculated using the cumulative binomial probability distribution for each sample size.

To extend these results to other sample sizes, Fig. 1 shows the probability of detecting 90% of the positive individuals (20% frequency) using a range of sample sizes for both random and patchy dispersion patterns (five patches) using the two sampling methods. When the dispersion pattern is random, both sampling methods give probabilities of detection close to the theoretical probabilities (Fig. 1A). When the dispersion pattern was patchy (Fig. 1B), random sampling again gave probabilities of detection close to the theoretical probabilities but systematic sampling performed poorly. Simulations using a 10% frequency of positive individuals gave similar results to those obtained for the 20% frequency.

TABLE 1: Effect of the degree of patchiness on the probability of detection of 90% of positive individuals at frequencies of 10 and 20% in the population. A sample size of 50 was used.

Frequency	Number of patches	Index of patchiness (<i>I</i>)	Probability of detection for random sampling	Probability of detection for systematic sampling
10%	Random	-0.002	0.56	0.56
	50	3.06	0.55	0.52
	30	4.16	0.59	0.40
	20	4.84	0.64	0.41
	10	5.70	0.55	0.29
	5	6.79	0.58	0.20
	2	7.27	0.58	0.18
20%	Random	-0.002	0.67	0.68
	50	2.03	0.69	0.64
	30	2.42	0.70	0.49
	20	2.83	0.76	0.48
	10	3.02	0.66	0.37
	5	3.09	0.71	0.30
	2	3.59	0.72	0.23

DISCUSSION AND CONCLUSIONS

Clearly, systematic sampling does not perform well when the frequency of the characteristic of interest in the population is low and patchily dispersed. This sampling method was more likely to underestimate the true frequency of positive individuals in the population. Furthermore, the sample sizes required to detect at least 90% of the individuals with a certain characteristic with high probability, are large. The simulations in the present study show that simple random sampling is not affected by the dispersion pattern of the characteristic in the population. Because the dispersion patterns of characteristics of interest are rarely known prior to sampling (and may even change over the period of sampling) systematic sampling should be avoided in favour of random sampling. For area or coordinate sampling, systematic sampling schemes are often chosen because of their simplicity and speed of completion, whereas random sampling usually requires the sampler to locate each sample unit from the (0, 0) coordinate, increasing the time required for completion.

To both simplify and speed up the implementation of a simple random area sample, Legg and Yeagan (1985) published a BASIC computer program that selects random coordinates but also calculates the shortest possible route to the next closest sample unit. The program is flexible enough to allow for the selection of coordinates from a variety of field sizes and sampling situations provided that the dimensions of the field or sampling frame are known beforehand. This program (rewritten in MS-DOS QBasic, and available from the senior author) runs on all versions of MS-DOS

and provides an output file that can be readily imported into a spreadsheet to generate a graphical representation of the sampling frame. Graphical representation allows the sample designer to check the sampling frame for poor distribution of random sample points that can occur when small sample sizes are used.

ACKNOWLEDGEMENTS

This research was supported by Lincoln University.

REFERENCES

- Basu, P.K., Lin, C.S. and Binns, M.R., 1977. A comparison of sampling methods for surveying alfalfa foliage diseases. *Can. J. Plant Sci.* 57: 1091-1097.
- Battacharyya, G.K. and Johnson, R.A., 1977. Statistical Concepts and Methods. John Wiley & Sons, Canada.
- Brun, L.O. and Suckling, D.M., 1992. Field selection for endosulfan resistance in coffee berry borer (Coleoptera: Scolytidae) in New Caledonia. *J. Econ. Entomol.* 85: 325-334.
- Krebs, C.J., 1989. Ecological Methodology. Harper Collins Publishers. N.Y.
- Landis, D.A. and Haas, M.J., 1992. Influence of landscape structure on abundance and within-field distribution of European corn borer (Lepidoptera: Pyralidae) larval parasitoids in Michigan. *Environ. Entomol.* 21: 409-416.
- Legg, D.E. and Yeagan, K.V., 1985. Method for random sampling insect populations. *J. Econ. Entomol.* 78: 1003-1008.
- Manly, B.F.J., 1992. The Design and Analysis of Research Studies. Cambridge University Press, Cambridge.
- Pedigo, L.P. and Buntin, G.D., 1994. Handbook of Sampling Methods for Arthropods in Agriculture. CRC Press, Boca Raton.