RASHI 1 (2) : 89-91 (2016)

A visit to a few research problems

Satyabrata Pal

Presently, Honorary Visiting Professor, Interdisciplinary Statistics Research Unit, ISI, Kolkata, India
Formerly, Dean, Post-Graduate Studies and Professor, BCKV, Mohanpur, Nadia, West Bengal, India

ABSTRACT

This article highlighted two major real life problems : The first problem is concerned with the determination of Economic Threshold Level (ETL) and Economic Injury Level (EIL) (Weersink et al, 1991) are two important parameters related to controlling the pest damage to crops. The existing method of determination of the ETL relies on some economic variables which are subject to considerable spatial and temporal variations. We have developed a new method of determination of ETL based on statistical approach. The method has been applied on some crops and the ETL of the corresponding crops have been obtained. The second problem pertains to the age-old problem of determination of optimum size and shape of plots (included as a Chapter in the celebrated book by Oscar Kempthorne published in the fifties of the twentieth century), the solution of which may be considered to be still open. Here we have obtained robust optimum plot sizes.

1. First Problem

The systematic procedure which has been developed by us requires weekly (usually) observations on incidences (count data) on the pest for which ETL is to be determined, needless to mention that real-life data are the prime selection from controlled experiments on research plots. Then these data are converted to per leaf count. The plot has to be maintained without spraying any pesticide throughout the year (crop life-span). Thus pests are subjected to their natural births, growths and deaths. The collected data on the weekly counts on incidences of the pests over the year (crop tenure on field) ensure the determination of the most appropriate probability distribution (probability density function) (among a number of competitive p.d.f.'s) by application of KS test.

The knowledge of the appropriate probability distribution enables us to obtain the probabilities of occurrences corresponding to different incidence values of the pest under consideration. In fact, a study on the incidence values (of pests) with low cumulative probabilities of occurrences is very much important to the scientists as incidence values (of pests) with high cumulative probabilities ensure enormous crop damage. The twin criteria, the economic injury level (EIL) or the "lowest population of pests that will cause economic damage", and the economic threshold level (ETL) or the "population density (number of pests) at which the control measures to be initiated in order to prevent an increasing pest population from reaching the economic injury level", are important. Indeed. The EIL is alarming as this level causes economic damage and ETL is the benchmark level, when reached, suggestive of determination of the control measures to be taken in order to prevent an increasing pest population from reaching the economic injury level (EIL).

The statistical approach determines that the cumulative probability range, "0.25 to 0.45", with .05 addition at the most on the upper extreme cumulative probability value, contains cumulative probability value corresponding to the ETL in most cases. Thus the ETL corresponding to a pest (with respect to a crop) can be identified by an examination of the above cumulative probability levels. In fact, ETL with respect to a particular pest (on a particular crop) is not fixed (undergoes slight variation), and thus ETL varies, though marginally, with the surrounding environment also.

On pest incidence data on white fly (*Singhiella palled*) and on black fly (*Aleurocanthus rugosa*) on betel vine (*Piper betle* L.) crop, application of K-S test identifies the Weibull distribution as the best fitting distribution for both cases. From two tables (not included here), one containing the occurrence-probabilities (cumulative probabilities) corresponding to the numbers of whiteflies and the other containing the occurrence-probabilities (cumulative probabilities) corresponding to of the numbers of blackflies respectively, it is found that the cumulative probability range, (0.25) to (0.45) corresponds to the range of number of pest occurrences (3-4) in both cases. In fact, the approximate intervals corresponding to the numbers, 3 and 4 are, (2.5-3.4) and (3.5-4.4) respectively. Thus the economic threshold level (ETL) values corresponding to the two pests, white fly and black fly (betelvine crop) lie in the range, (3-4). The method has worked well for a few other pests on different crops which we have examined. However, our method needs to be applied on many other pests to determine its universal applicability.

Note : This article is intended to be a brief overview on some of the research problems (real-life) with which I am involved at the present moment with my erstwhile students of BCKV, my present research scholar, and my colleague at ISI.

E-mail : spbr18@gmail.com

2. Second Problem

The aspect of developing appropriate procedures leading to the determination of optimum plot size in field experiments can be found in the papers, namely, Smith (1938), Zhang *et al*. (1990, 1994), Bhatti *et al*. (1991), Fagroud and Meirvenne (2002), to name a few. Pal *et al.* (2007) presented a new approach to determine the optimum plot size in the context of the evidence that even with Fisherian blocking correlation among the residuals is not eliminated in field experimental data. Subsequently, in Pal *et al.* (2015), theoretical exploration of the effect of different plot sizes and shapes in discovering the point (actually the minimum radius of curvature of the variogram at that point beyond which the theoretical variogram assumes almost stationary values with further increase in lags) has been pursued in case of the most commonly employed model (incorporating correlation structure) assumed to represent real-life data situations (uniformity trial or designed experiments, RBD/LSD) precisely.

To define a variogram, let $\{Y(s): s \in D_s\}$ be a real valued spatial process defined on a domain D_s of the 2-dimensional Euclidean space R^2, and it is supposed that the variance of the difference of the values of the variable at s_1 and s_2 (displaced h-apart, *i.e.*, $s_1 = s$, and $s_2 = s + h$) vary in a way that depends only on $s_1 - s_2 = h$, $h > 0$. Symbolically, Var $[Y(s + h) - Y(s)] = 2\gamma(Y(h)) = 2\gamma(h)$, for all $(s, s + h) \in D_s$, the variogram, however, satisfies the conditional-non-positive-definiteness condition. $\gamma(h)$ is called the semi-variogram. The quantity $2\gamma(h)$ being a function of the difference between the spatial locations s and s + h, is called the stationary variogram. When $2\gamma(h)$ becomes independent of s, and is a function of $||h||$ only, for $h = (h_1, h_2)$ R^2, $|| h || = (h_1^2 + h_2^2)^{1/2}$, the variogram is said to be isotropic, otherwise, it is said to be anisotropic. For further reading, the paper by Mathron (1963) and book Cressie and Wikle (2011) may be consulted.

The Uniformity trial data Y (s) on a spatial location s is modelled as:

$Y(s) = \mu + e(s)$, $V(Y(s)) = (V(e(s)) = \sigma^2$; Cov $(Y(s), Y(s + h)) = \text{Cov}(e(s), e(s + h)) = \rho^{||h||}\sigma^{2"}$…… (1);

OR,

$Y(s) = \mu + e(s)$, $V(Y(s)) = (V(e(s)) = \sigma^2$; Cov $(Y(s), Y(s + h)) = \text{Cov}(e(s), e(s + h)) = \rho^{|h_1|+|h_2|}\ \sigma^2$ …….. (2).

Model (2) is relatively simpler (in form) than the Model (1) but also widely suitable and applicable in real-life data originated from field experiments. Ample applications of this model abound the literature to represent spatial situations pertaining to the data generated from field experiments.

Variogram of the residuals are to be modelled in case of data collected from the designed experiments (RBD, LSD, *etc.*) adopting the method delineated in the paper. It is to be noted that model (1) induces more variability in $2\gamma(h)$ values in comparison to model (2)**.**

The expressions of the theoretical variograms, $2\gamma(h)$ (under the above models (1) and (2)) have been obtained for plot sizes, l x k (l = 2, 3, ……; k = 2, 3, ……), *i.e.*, 2 x 2, 2 x 3, 2 x 4, 2 x 5, 2 x 6, 2 x 7, 2 x 8, 3 x 3, 3 x 4, 3 x 5, 3 X 6, 3 x 7, 3 x 8, 4 x 4, 4 x 5, 4 x 6, 4 x 7, 5x5, 5x6, *etc.*, respectively (area of plots being less than or equal to 30 squared units).

Expressions of $2\gamma(h)$ values (for plot sizes, 2x2 and 2x5) are presented in case of Model (2) as examples [expressions of $2\gamma(h)$ values are not given here (available in Pal, *et al.* (2015)]:

Plot size: 2x2:

$$2\gamma(h) = \sigma^2[8 + 4(2\rho^2 + 4\rho) - 2(2\rho^{2h+2} + 6\rho^{2h+1} + 6\rho^{2h} + 2\rho^{2h-1})]$$

Plot size: 2x5:

$$2\gamma(h) = \sigma^2[20 + 4(2\rho^5 + 6\rho^4 + 10\rho^3 + 14\rho^2 + 13\rho) - 2(2\rho^{5h+5} + 6\rho^{5h+4} + 10\rho^{5h+3} + 14\rho^{5h+2} + 18\rho^{5h+1} + 18\rho^{5h} + 14\rho^{5h-1} + 10\rho^{5h-2} + 6\rho^{5h-3} + 2\rho^{5h-4})]$$

The alternative robust optimum plot sizes are determined as, 2x5, 2x6, 3x5, 3x6, 5x5 under the above model structure, (1), the values of the intra-class correlation (ρ) of the first order being, $\rho = 0.1$, $\rho = 0.2$, $\rho = 0.3$, $\rho = 0.4$, and $\rho = 0.5$, respectively.

The alternative robust optimum plot sizes are determined as, 2x5, 2x6, 2x7, 3x5, 3x6, 3x7, 4x5, 4x6, 4x7, and 5x5 under the above model structure, (2), the values of the intra-class correlation (ρ) of the first order being, $\rho = 0.1$, $\rho = 0.2$, $\rho = 0.3$, $\rho = 0.4$, and $\rho = 0.5$, respectively.

The higher values of ρ ($\rho > 0.5$) are not evidenced in real-life data from field experiments. In order to take care of spatial heterogeneity in two directions, the choices in regard to the plot sizes have been restricted to plot sizes, 2 x 2, and higher. It is to be noted that

expressions for 2 Y (h) in case of one-dimensional variants (which do not reflect the scenario of spatial heterogeneity satisfactorily, as, long and narrow plot sizes may induce more heterogeneity).

REFERENCES :

First problem :

Pal, S., Ghosh A, and Dhar, T. Dhar, (2013): On Determination of ETL – A Distributional Approach, *Biometrical Letters*, **50** : 107-16.

Weersink, T. A., Deen, W. and Weaver, S. (1991): Defining and Measuring Economic Threshold Levels. *Canadian J. Agric. Econ*. **39** : 619-25.

Second problem :

Bhatti, A.U., Mulla, D.J., Koehler, F.E. and Gurmani, A.H. (1991): Identifying and removing spatial correlation from yield experiments. *Soil Sci. Soc. Am. J.* **55**: 1523-28.

Cressie N., Wikle C.K. (2011): *Statistics for Spatio-Temporal Data*. A John Wiley & Sons. Inc publication.

Faground M., Meirvenne M. Van (2002): Accounting for Soil Spatial Autocorrelation in the design of experimental trials. *Soil Sci. Soc. Am. J.* **66**: 1134-42.

Matheron G. (1963): Principles of geostatistics. *Economic geology*, **58**: 1246-66.

Pal S., Basak S, Kageyama, and Sanpei (2007): On determination of optimum size and shape of plots in field trials. *Biometrical Letters*. **44**: 23-31

Pal S, Mandal, G, and Dihidar K., (2015): Determination of Robust optimum Plot Size and Shap. *Biometrical Letters.* **52** : 13-22.

Smith H.F. (1938): An empirical law describing heterogeneity in the yields of agricultural crops. *J. Agric. Sci., Cambridge*. **28**: 1-29.

Zhang R, Warrick A.W., Myers D.E. (1990): Variance as a function of sample support size". *Math. Geol.* **22** : 107-21.

Zhang, R, Warrick, A.W., and Myers D.E. (1994): Heterogeneity, plot shape effect and optimum plot size. *Geoderma*, **62**: 183-97.