THE NUMBER OF RED ROT LESIONS ON THE MID-RIB OF SUGARCANE LEAVES

BY JAGDISH NARAIN SRIVASTVA

Indian Institute of Sugarcane Research, Lucknow

1. INTRODUCTION

LESIONS are caused on the mid-ribs of sugarcane leaves by primary infection when the crop is young and later by secondary infection of red rot on the sugarcane crop. This formation of a lesion starts as reddish dots on the upper surface of the mid-rib. Later the central regions of these reddish portions turn straw coloured studded with small black dot like structures known as acervuli. The central straw coloured portion is surrounded by reddish margin at the edges.

A leaf may have no lesion, in which case it may be said to be uninfected. The infected leaves, on the other hand, may have one, two, three, etc., lesions. It is the purpose of this paper to study the distribution of the infected leaves according to the number of lesions they possess. Formula for estimating the parameter occurring in this distribution has also been worked out. Confidence interval for the estimate of the parameter in case of large samples has been determined, and the size of the sample required for a sufficiently accurate estimation of the parameter has been discussed. Also the distribution of the maximum likelihood estimate of the parameter has been found out.

2. DISTRIBUTION OF LEAVES ACCORDING TO THE NUMBER OF RED ROT LESIONS THEY POSSESS

An attempt can be made to find out theoretically what this distribution would be. Let us imagine the leaves to be surrounded by an infectious atmosphere in such a way that all of them have got equal chance of being attacked. Let q represent* the intensity of infection, meaning thereby that if there are m leaves, the portion mq of them will become infected, and will develop a lesion and the portion $m(1-q)$ will remain uninfected. In the same way a portion mq^2 will develop one more lesion out of the mq leaves containing one lesion. A portion mq^3 out of these will develop a third lesion and out of these mq^3 leaves, a portion mq^4 shall develop a fourth one and so on. The basic prob-

* In fact q depends both upon the intensity of the disease and the susceptibility of the leaves.

ability law of the distribution of the number of red rot lesions on the mid-rib of sugarcane leaves, may, therefore be described as

$$P = c \Sigma q^x \ldots\ldots, \tag{1}$$

where q is the probability of getting a leaf infected and having one lesion, c is a constant, x denotes the number of lesions in a leaf.

The probability law assumes the independence of the formation of successive number of lesions after one lesion has been developed. In case the probability law (1) is being used to specify the population including both infected and uninfected leaves, x takes the values $0, 1, 2, \ldots,$ and c can be seen to be equal to $1/(1 - q)$. But if we use the probability law to describe only the truncated population excluding the value $x = 0$, c can be easily evaluated as $(1 - q)\, q^{-1}$.

Whether the above probability law agrees with observations can be verified from mycological data collected during November 1955, and November 1956. In 1956 November, a sample of 816 leaves was drawn randomly from 1/7 acre field of sugarcane at I.I.S.R., Lucknow. The whole of the sample was drawn during the same day, and during that very day each leaf was scrutinised with respect to the number of red rot lesions it contained. It was found that out of these 816 leaves examined, 497 were unaffected, and 319 were infected. A sample of 100 infected leaves was drawn in exactly the same manner as above in November 1955. Data of both years in detail, is being reproduced in Table I.

To examine the goodness of fit of the (discrete) distribution (1) in the above data, χ^2 was calculated firstly for the year 1956, in which case the population is complete, and the case $x = 0$ is included. The value of q put in the formula (1) for calculating expected y_x is the maximum likelihood estimate of q and is given by

$$q = \frac{\bar{x}}{\bar{x} + 1}$$

The value of χ^2 obtained is 16·21 for 5 *d.f.*, the probability for obtaining which is less than 0·01. It, therefore, appears at first sight that the distribution (1) cannot represent fully the said observed values.

However let us consider the second case, *viz.*, the truncated frequency distribution given by

$$y_x = n\,(1 - p)p^{x-1}, \tag{2}$$

for value of $x = 1, 2, 3, \ldots.$ Here n refers to the number of infected leaves only, p is the chance of getting one lesion, and y_x the number of leaves containing x lesions. This distribution will represent all the figures in column (3), Table I, except the first figure, *viz.*, the number of unaffected leaves, and all the figures in column (2).

TABLE I

Number of Sugarcane Leaves according to the Number of Red Rot Lesions they possess

No. of lesions in a leaf (x)	No. of leaves possessing x lesions (y_x)	
	(1955)	(1956)
0	Figure not available	497
1	62	158
2	26	81
3	8	46
4	3	17
5	1	11
6	Nil	5
7	Nil	1
More than 7	Nil	Nil
TOTAL ..	100	816

It is found that the frequency function (2) gives a good fit to the data to which it refers. Using the maximum likelihood estimate of p (see below), this curve was fitted to represent the distribution of infected leaves in each of the two years. Values of χ^2 equal to 3·057 at 4 *d.f.* and 0·613 at 2 *d.f.*, were obtained respectively for the years 1956 and 1955, which correspond respectively to probabilities of 56% and 74%. It, therefore, appears that the distribution of the number of infected leaves (y_x) according to the number of red rot lesions they possess is given approximately by the equation (2).

The above fact is also supported by mycological considerations. It appears that the value of p to be used for finding out the proportion -- infected or uninfected leaves should not be the same as that to be used for determining the number of lesions in leaves once they have got infected, since in the former case some more factors concerning the susceptibility of the leaf come into play.

It has been found that the spores bearing length on a leaf is much less for leaves with more than one lesion than for one lesion leaves. Since the spores bearing length on a leaf is directly responsible for the further propagation of the disease, it becomes important to estimate the proportion of leaves having one, two, etc., lesions and hence the value of p.

It is proposed to elaborate the above points in another paper.

3. The Maximum Likelihood Estimate of p

The likelihood of the sample $(x_1 \ldots . x_n)$ is given by

$$\phi = (1-p)^n p^{n\bar{x}-n}, \tag{3}$$

so that

$$\partial \log \phi / \partial p = 0, \text{ gives } \hat{p} = (\bar{x}-1)/\bar{x} \tag{4}$$

as the maximum likelihood estimate of p, and where x denotes the mean number of lesions per leaf in the sample.

4. Confidence Interval for p and Size of the Sample Required for a Sufficiently Accurate Estimate of p

By applying the Central Limit Theorem to the variate $\partial \log \phi / \partial p$, we have

$$\frac{\frac{\partial \log \phi}{\partial p}}{\sqrt{V\left(\frac{\partial \log \phi}{\partial p}\right)}} \sim N(0, 1)$$

in case the size of the sample is large.

We have

$$\frac{\partial \log \phi}{\partial p} = \frac{n}{p}\left\{\bar{x} - \frac{1}{(1-p)}\right\}$$

and hence

$$V\left(\frac{\partial \log \phi}{\partial p}\right) = \frac{n}{p(1-p)^2}, \tag{5}$$

since

$$V(\bar{x}) = \frac{p}{n(1-p)^2},$$

which can be proved as below:

We have

$$E(x) = \sum_{1}^{\infty} x(1-p)p^{x-1}$$

$$= (1-p)(1-p)^{-2} = (1-p)^{-1}$$

$$E(x^2) = \sum_{1}^{\infty} x^2(1-p)p^{x-1}$$

$$= (1-p)\left\{1 + \sum_{x=1}^{\infty}(x+1)^2 p^x\right\}$$

$$= (1-p) + pE(x^2) + 2pE(x) + p.$$

so that

$$E(x^2) = \frac{(1+p)}{(1-p)^2},$$

and therefore

$$V(x) = \left[\frac{(1+p)}{(1-p)^2}\right] - \left[\frac{1}{(1-p)^2}\right] = \frac{p}{(1-p)^2}.$$

Since $x_1, x_2, \ldots . x_n$ have been drawn independently,

$$V(\bar{x}) = \frac{p}{n(1-p)^2}.$$

Thus we have for large n,

$$\frac{\frac{n}{p}\left\{\left(\bar{x} - \frac{1}{1-p}\right)\right\}}{\frac{\sqrt{n}}{\sqrt{p(1-p)^2}}} \sim N(0, 1).$$

If we define d_α as

$$\int_{-d_\alpha}^{d_\alpha} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \alpha, \tag{6}$$

where α is the confidence probability, then we have on simplification,

$$\text{Prob.} \left\{ \left| \frac{\sqrt{n}}{\sqrt{p}} \{(1-p)\bar{x} - 1\} \right| \leqslant d_\alpha \right\} = \alpha \tag{7}$$

Thus if the lower and upper ends of the confidence interval of d are denoted by $\hat{p}_1$ and $\hat{p}_2$, these will be given by the roots of the equation,

$$\frac{n}{p}\{(1-p)\bar{x} - 1\}^2 = d_\alpha^2 \tag{8}$$

In fact we shall have from (8),

$$\hat{p}_1 + \hat{p}_2 = \left(2 - \frac{2}{\bar{x}}\right) + \frac{d_a^2}{n\bar{x}^2}$$

and

$$\hat{p}_1\hat{p}_2 = \frac{(\bar{x} - 1)^2}{\bar{x}^2} \tag{9}$$

Also from (9), we can find the length of the confidence interval as

$$(\hat{p}_2 - \hat{p}_1)^2 = \frac{d_a^2}{n^2\bar{x}^4}\{4n\bar{x}(\bar{x} - 1) + d_a^2\} \tag{10}$$

If we desire to choose n such that the length of the confidence interval be $1/k$th of the estimate p, we shall have the relation

$$\hat{p}^2 = k^2(\hat{p}_2 - \hat{p}_1)^2, \tag{11}$$

which in conjunction with equations (4) and (10) gives

$$n = kd_a^2 \left\{\frac{2k + \sqrt{4k^2 + 1}}{\bar{x}(\bar{x} - 1)}\right\}$$

which is nearly the same as

$$n \doteq \frac{4k^2 d_a^2}{\bar{x}(\bar{x} - 1)} \tag{12}$$

The equation (12) is important for determining the sample size required for attaining a particular degree of accuracy in the estimate, with a particular probability. As such the values of n have been calculated from equation (12) for various values of k, a and $\bar{x}$, and are given in Table II. As will be seen from equation (12), some preliminary idea of the value of $\bar{x}$ is necessary. This can be obtained, among other means, by having a small preliminary sample.

For example, the value of $\bar{x}$ for infected leaves in the year 1956 was 1·93, which means a value of p of about 48 per cent. A sample of size about 200 was necessary, therefore, in order that the true value of p might, with 95 per cent. probability, lie on either side within a range of 10 per cent. ($k = 5$) of the estimated value of p.

5. Distribution of $\hat{p}$

One may be interested in knowing the distribution of $\hat{p}$ since it may be needed in many applications of the curve (2). For this purpose we shall first find the distribution of $\bar{x}$, the arithmetic mean of a sample of size n drawn from the population given by (2).

TABLE II

Showing the Sample Size required for Attaining Various Degrees of Accuracy in the Estimate of p

1	2	3	4
p	$\bar{x}$	n	n
Corresponding to $\bar{x}$ in column 2	from preliminary estimate	$k = 10$ $\alpha = 0 \cdot 95$	$k = 5$ $\alpha = 0 \cdot 95$
0·05	1·0526	27,750	6,940
0·10	1·1111	12,450	3,110
0·20	1·2500	4,900	1,225
0·30	1·4286	2,500	625
0·40	1·6667	1,400	350
0·50	2·0000	750	190
0·60	2·5000	400	100
0·70	3·3333	200	50
0·80	5·0000	75	20

Note.—Values of p obtained with the help of (4) from the value $\bar{x}$ calculated from the sample have also been given for having an idea of the same.

Let the sample be $x_1, x_2, \ldots x_n$, where each x_i can be 1, 2, 3.... The probability of drawing the sample is given by (3).

$$\text{Prob.}\,(x_1, x_2, \ldots x_n) = (1 - p)^n\, p^{n\bar{x}-n}$$

Since an x_i can be any one of the integers 1, 2, 3...., it is clear that $\bar{x}$ can have values only of the form $\bar{x} = 1 + k/n$ where $k = 0, 1, 2, \ldots$ To determine the distribution of $\bar{x}$ therefore, we have to find the probability that $\bar{x} = 1 + k/n$, for $k = 0, 1, 2 \ldots$.
or in other words,

$$x_1 + x_2 + \ldots + x_n = n + k \tag{13}$$

Thus if ϵ_k represents the number of sets of values of $(x_1, x_2 \ldots . x_n)$ such that the condition (13) holds we shall have

$$\text{prob.}\left(\bar{x} = 1 + \frac{k}{n}\right) = \epsilon_k (1-p)^n p^{n\bar{x}-n} \tag{14}$$

which shall represent the probability distribution of $\bar{x}$.

To evaluate ϵ_k, we note that it is equal to the number of ways in which $n + k$ objects can be placed at n places such that each place contains at least one object and each object is placed in some one out of the n places. Consequently it can be verified that

$$\epsilon_k = {}^{n+k-1}c_k = \frac{(n+k-1)\,!}{(n-1)\,!\;k\,!} \tag{15}$$

Combining (14) and (15), we can express the distribution of $\bar{x}$ as

$$\text{prob.}\left(\bar{x} = 1 + \frac{k}{n}\right) = (1-p)^n\ {}^{n+k-1}c_k\, p^k,\ k = 0, 1, 2, \ldots \tag{16}$$

To derive the distribution of p, we observe that each value of $\bar{x}$ corresponds to one and only one value of p and *vice versa.* When $\bar{x} = 1 + k/n$, we have

$$\hat{p} = \left(1 + \frac{n}{k}\right)^{-1},\ k = 0, 1, 2 \ldots . \tag{17}$$

Hence the distribution of $\hat{p}$ is given by

$$\text{prob.}\left\{\hat{p} = \left(1 + \frac{n}{k}\right)^{-1}\right\} = (1-p)^n\ {}^{n+k-1}c_k p^k,$$

$$\text{for } k = 0, 1, 2, \ldots . \tag{18}$$

or in a slightly different form, by

$$\text{prob.}\,(\hat{p}) = \frac{(1-p)^n}{(n-1)!}\ \frac{\left(\dfrac{n}{1-\hat{p}} - 1\right)!}{\left(\dfrac{n\hat{p}}{1-\hat{p}}\right)!}\ p^{n\hat{p}/1-\hat{p}} \tag{19}$$

The distribution of $\bar{x}$ and hence that of p can also be obtained by using characteristic functions. For, since the characteristic function $H_x(t)$ of the distribution (2) is given by

$$\begin{aligned} \text{H}_x(t) &= \text{E}\,(e^{ixt}) \\ &= (1-p)\, e^{it}\, (1-pe^{it})^{-1}, \end{aligned} \tag{20}$$

so that that of the distribution of mean $\bar{x}$ shall be given by,

$$H_{\bar{x}}(t) = (1 - p)^n e^{it} (1 - pe^{it/n})^{-n}, \tag{21}$$

which it can be seen is the C.F. of the distribution (16).

6. Summary

It is seen that the distribution of the number of sugarcane leaves according to the number of red rot lesions in them is non-normal, the curve (2) giving a good approximation to it. For estimation of p, the parameter occurring in the distribution, with a particular degree of accuracy, a table giving the sample size required has been constructed. The distribution of the maximum likelihood estimate of p has also been derived. Further investigations in this line are in progress and shall be presented in another paper.

7. Acknowledgement

I am deeply indebted to Mr. S. A. Rafay, Mycologist, Indian Institute of Sugarcane Research, Lucknow, who kindled my interest in this work and provided me with necessary data. I am also highly grateful to Dr. B. K. Mukerji, Director, Indian Institute of Sugarcane Research, Lucknow, for his kind encouragement and advice.