ANOTHER TEST FOR HETEROGENEITY OF HOST RESISTANCE IN DILUTION ASSAYS

By P. A. P. MORAN

Australian National University, Canberra, A.C.T.

Suppose that λ is the average density per unit volume in a suspension of infective particles such as virus particles. To estimate λ the usual method is to make up a series of inocula of various dilutions containing expected numbers of particles $\ldots\lambda_{i-1}, \lambda_i, \lambda_{i+1},\ldots$ which are known multiples of λ. Each of these is then tested by inoculation in a host such as an egg. We consider only the case where the dilution series is twofold ($\lambda_i = \lambda 2^i$, say) and the same number of eggs, N, is tested at each dilution. Then if we are sure that an egg is infected if and only if the inoculum contains at least one infective particle, the probability that the egg remains sterile is $P_i = \exp\{-\lambda 2^i\}$. If each particle is not certainly infective but has a probability p of infecting the egg, the probability of sterility of an egg chosen at random is $\exp\{-\lambda p 2^i\}$ provided that p does not vary from egg to egg. It is then only possible to estimate λp from the results.

If, however, p varies from egg to egg with a probability distribution $f(p)$, the probability of an egg, chosen at random, being sterile is

$$\int_0^1 e^{-\lambda p 2^i} f(p)\, dp$$

and when plotted against i, this gives a flatter curve. Thus any test of the goodness-of-fit of the original hypothesis provides us with a method of testing for variation in host resistance.

Instead of fitting by maximum likelihood and then using χ^2 for this purpose, a more effective test has been proposed (Moran (1954*a*, *b*)). Provided that the series is sufficiently long to range from almost certainly sterile to almost certainly fertile levels we calculate a quantity $T = \sum_m f_m(N - f_m)$. Here f_m is the number of fertile eggs at the dilution level 2^m and N is the number of eggs tested at each level. The mean and variance of T have been calculated, thus enabling a rapid test to be made on the assumption that T is approximately normally distributed. This test has the advantage of being very much faster than the χ^2-test and also appears, in most cases, to be substantially more powerful (Armitage & Spicer (1956)), since it is so constructed that T is large for one particular kind of divergence from the theoretically expected numbers, namely, that in the direction of a flattening of the graph of P_i against i.

Another method is based on the Spearman–Kärber approach but is shown by Armitage & Spicer (1956) to be less efficient than the T-test. Fazekas de St Groth (1955) has also examined the practical use of these tests.

One disadvantage of the T-test is that it gives equal significance to two such series as

...000123345555...

and

...000123355545...,

https://doi.org/10.1017/S0022172400037815 Published online by Cambridge University Press

whereas clearly the latter indicates a stronger divergence from the null hypothesis. In view of this another test is here proposed.

As a test criterion we take the distance, D say, between the first level at which at least one egg is sterile and the last level at which at least one egg is fertile. Thus for $N=5$ and the series

$$\ldots 000123354555\ldots$$

D is equal to 5. We calculate the probability distribution of D on the assumption that $P_i = \exp\{-\lambda 2^i\}$. Write $Q_i = 1 - P_i$. Then

$$\text{prob}\{D=s\} = \sum_{m=-\infty}^{\infty} \ldots P^N_{m-2} P^N_{m-1} (1-P^N_m)(1-Q^N_{m+s}) Q^N_{m+s+1} Q^N_{m+s+2} \cdots,$$

where $s = 1, 2, \ldots$. The infinite products involved are easily seen to be convergent. This probability clearly depends on λ, but since it is a periodic function of $\log_2 \lambda$ with period 1 we may hope, in analogy with the previously given theory of the T-test, that it will vary only very slightly with λ. Write

$$A_s = \ldots P_{s-1} P_s, \quad B_s = Q_s Q_{s+1} \cdots$$

and put

$$T_{N,s} = \sum_{m=-\infty}^{\infty} A^N_m B^N_{m+s}.$$

Notice that

$$A_m = \exp\{-\lambda \sum_{r=m}^{\infty} 2^r\} = \exp\{-\lambda 2^{m+1}\} = P_{m+1}.$$

Then

$$\text{prob}\{D=s\} = \sum_{m=-\infty}^{\infty} \{A^N_{m-1} - A^N_m\}\{B^N_{m+s+1} - B^N_{m+s}\}$$

$$= T_{N,s+2} - 2T_{N,s+1} + T_{N,s}.$$

Besides $D = 1, 2\ldots$ there are two other possibilities; first that the series is of the form $\ldots 000aNN\ldots$, where $0 < a < N$, and secondly, that it is of the form $\ldots 000NNN\ldots$. These are easily seen to have probabilities $T_{N,2} - 2T_{N,1}$, and $T_{N,1}$, respectively.

If $T_{N,s}$ is averaged over the possible values of $\log_2 \lambda$ in the interval (0, 1) we obtain the integral

$$\int_{-\infty}^{\infty} \exp\{-N 2^{x+1}\} \prod_{r=0}^{\infty} \{1 - \exp\{-2^{x+s+r}\}\}^N dx,$$

which unfortunately does not appear to be expressible in terms of elementary functions. We must therefore evaluate $T_{N,s}$ numerically and explore the extent of its variation (or rather, of the variation of prob $\{D=s\}$) with λ. This has been done for $N=4$ and $N=10$, and also the probabilities that $D=s$ found for $N=5$, 8, and one particular value of λ. The results are presented in Table 1 in the form of the probabilities of the tails of the distribution in its statistically interesting part.

This table was constructed in the following way. For $N=4$, 10 five values of λ were chosen so that one of the probabilities, say, was equal to 0·999000, 0·998800, ...0·998200, respectively (0·001000, 0·001201, 0·001401, 0·001601 and 0·001802) and the values of $T_{N,s}$ found to six places of decimals. The probabilities of the ordinates and tails of the distribution were then calculated from these, also to six decimal places and used for the second and fifth columns of the above table. It was found

https://doi.org/10.1017/S0022172400037815 Published online by Cambridge University Press

that for $N=4$ and $s \geqslant 5$, the maximum error due to variation in λ is not greater than 0·000036, whilst for $N=10$ and $s=7$ the maximum error is not greater than 0·000148. This covers the statistically interesting parts of the tails. These errors being very small the values for $N=5$ and $N=8$ were calculated solely for $P_0=0{\cdot}999000$ and rounded off to four places. In all cases it was found that sufficiently far along the tail the probabilities of the ordinates were decreasing in such a way that their successive ratios were of the form $\frac{1}{2}+K2^{-r}$, where K is a constant. By extrapolation the calculation of the tail for the larger values of s was thus very

Table 1. *Probabilities that* $D \geqslant s$

s	$N=4$	$N=5$	$N=8$	$N=10$
5	0·3131	0·3955	0·5996	0·6999
6	0·1755	0·2289	0·3804	0·4697
7	0·0932	0·1239	0·2177	0·2787
8	0·0481	0·0645	0·1170	0·1530
9	0·0244	0·0330	0·0607	0·0803
10	0·0123	0·0167	0·0309	0·0412
11	0·0062	0·0084	0·0156	0·0208
12	0·0031	0·0042	0·0078	0·0105
13	0·0016	0·0021	0·0039	0·0053
14	0·0008	0·0011	0·0020	0·0026
15	0·0004	0·0005	0·0010	0·0013
16	0·0002	0·0003	0·0005	0·0007

easy and the whole calculation received a partial check by verifying that the sum of the probabilities, together with the remaining tail, and the probabilities of the two additional cases mentioned above, was unity.

The probability distributions for $N=6$, 7 and 9 can be found from the above by linear interpolation with an accuracy which is quite adequate as is shown by comparing the values for $N=8$ with a linear interpolate from $N=5$ and $N=10$. Thus for $s=7$, 8 and 9 we get by interpolation values 0·2168, 0·1176 and 0·0614 which differ from the values given in the table by $-0{\cdot}0009$, 0·0006 and 0·0007, respectively. Thus interpolation for $N=6$, 7 and 9 will be quite accurate.

As an example of the use of the test consider the observed sequence 0, 0, 2, 1, 1, 3, 4, 5, 4, 5, 5 ($N=5$). Here $D=6$ and the probability of as large or larger a value of D is 0·2289.

The power of this test would be difficult to determine in any general situation although the change in the distribution of D for particular numerically specified alternatives of the kind given by Armitage & Spicer (1956) could be found, and this would be worth doing. Furthermore, the correlation coefficient between T and D could be found numerically without much difficulty. A small amount of experience suggests that they are fairly closely related. It seems plausible to suggest that the present test is less efficient than T since it does not, in a sense, use all the available information. However, some kinds of alternative hypotheses often seem in practice to give rise to outliers of fertile eggs at high dilutions and in this type of situation the present test may be useful. Armitage & Spicer rightly point out that all known tests of this kind may easily fail to pick up considerable

heterogeneity in host resistance for the usual values of N. This is partly due to the fact that we do not know λ and partly because the estimator of a binomial probability has little intrinsic accuracy. Probably the ideal method would be to compare two sets of inoculations one of which was known for certain to contain exactly one potentially infective particle, and the other exactly two.

This test could also be extended to other dilution series. A fourfold dilution series would probably result in a distribution which was reasonably invariant with λ but a tenfold dilution series would almost certainly be unsatisfactory. Moreover, for such series the smaller number of ordinates in the main part of the distribution impairs the test.

REFERENCES

ARMITAGE, P. & SPICER, C. C. (1956). The detection of variation in host susceptibility in dilution counting experiments. *J. Hyg., Camb.*, **54**, 401.

FAZEKAS DE ST GROTH, S. (1955). Production of non-infective particles among influenza viruses; do changes in virulence accompany the von Magnus phenomenon? *J. Hyg., Camb.*, **53**, 276.

MORAN, P. A. P. (1954*a*). The dilution assay of viruses. I. *J. Hyg., Camb.*, **52**, 189.

MORAN, P. A. P. (1954*b*). The dilution assay of viruses. II. *J. Hyg., Camb.*, **52**, 444.

(*MS. received for publication* 1. I. 58)