Distribution Pattern of the Duplication of Discoveries

PRICE[1] conjectured that a Poisson distribution might a good fit for the pattern of incidence of multiple d coveries. We have fitted a modified Poisson distribution the frequency of duplication of discovery of antibio

We have considered the discovery of an antibiotic have been duplication if it was identified with anoth antibiotic and so reported in published documen irrespective of whether the compounds were produced one and the same or different species of microorganism We used the data reported earlier[2] on the frequency a duplication of discovery of antibiotics during 1907–19 after correcting a few minor discrepancies. As there w practically no duplication of discovery of antibiotics fo bacteria, we have dealt only with the pattern relating two groups: (1) fungi, algae and lichens, and (2) actinomycetales.

A normal Poisson distribution did not fit the distribution pattern of duplications, but a modified version[3] did For such a distribution, the p.d.f. is

$$\binom{c}{c+1}^p \left\{\frac{p(p+1)(p+2)\ldots(p+(r-1))}{(r-1)!\ \ (r+1)^r}\right\} \quad (1$$

where p and c are constants to be estimated using th following relations

$$\text{mean of } X = p/c \quad (2$$

$$\text{var}(X) = p/c + p/c^2 \quad (3$$

Using equations (2) and (3), p and c were estimated separately from the data on antibiotics derived from the two groups of organisms and for the pooled data. Using formula (1), the expected frequencies were computed A χ^2 test was done. To fit the distribution, the upper tail of the distribution from duplication frequency 8 and above was omitted. Table 1 gives details of the goodness of fit. In all three cases, the goodness of fit was confirmed.

Table 1. OBSERVED AND EXPECTED FREQUENCIES OF DUPLICATION, AND TEST OF GOODNESS OF FIT

No. of times duplicated	No. of antibiotics from fungi, algae and lichens		No. of antibiotics from Actinomycetales		Total	
	Observed	Expected*	Observed	Expected*	Observed	Expected
0	371	364·2	1,018	1,011·6	1,389	1,389·5
1	39	44·3	79	81·0	118	121·5
2	12	16·8	20	29·6	32	44·2
3	7	6·9	15	13·3	22	20·2
4	2†	3·3	7	6·7	9	10·1
5	4†	1·6	7†	3·5	11	5·4
6	2†	0·8	4†	1·9	6†	2·9
7	1†	0·4	1†	1·1	2†	1·7
	2·98		8·23		12·15	
χ^2_{cal}	9·49 for 4 df		12·59 for 6 df		12·59 for 6 df	

* Value corrected to first decimal place.
† Grouped data used.

There thus seems to be predictable regularity in duplications, and Price's conjecture about the pattern of distribution of multiple discoveries is confirmed for antibiotics, even in cases where there are no duplications.

A. NEELAMEGHAN
V. V. BUCHE
B. S. S. GUPTA

Documentation Research and Training Centre,
Indian Statistical Institute,
112 Cross Road 11,
Bangalore 3, India.

Received December 8, 1969; revised July 7, 1970.

Price, D. De S., *Little Science, Big Science*, 66 (Columbia University Press, New York, 1963).
Neelameghan, A., *Library Science with a Slant to Documentation*, **5**, 251 (1968).
Kendall, M. G., and Stuart, A., *Advanced Theory of Statistics*, second ed., **1** (Hafner, New York, 1958).