236

THE SIGNIFICANCE OF DEVIATIONS FROM EXPECTATION IN A POISSON SERIES

R. A. Fisher

University of Cambridge

In a series of observations of whole numbers in which the number r is observed a_r times, we may set

$$\underset{r}{S}(a_r) = A,$$

$$\underset{r}{S}(ra_r) = B,$$

then the ratio

$$\bar{r} = B/A$$

is a sufficient estimate of m, where m is the parameter of the Poisson Series with expectations

$$m_r = A\, e^{-m} \frac{m^r}{r!}\,.$$

If $\bar{r}$ is small and the series short, evidence of deviation may come chiefly or wholly from frequencies with low expectation, so that the measure of general discrepancy

$$\chi^2 = \underset{r}{S}(a_r - m_r)^2/m_r$$

will have a sampling distribution not accurately given by the usual table of χ^2. Equally, the special test for discrepancy of the variance, the index of dispersion

$$\chi^2 = \underset{r}{S}\{a_r(r - \bar{r})^2\}/\bar{r}$$

for $A - 1$ degrees of freedom, will be suspect owing to the low estimated expectation, $\bar{r}$, in each of the A cells.

Biometrics, 6: 17-24, (1950).

PARTICULAR EXAMPLE

These remarks may be illustrated by the particular example

r	frequency
0	124
1	12
2	2
3	2
Total	140

The mean, $\bar{r}$, is only 11/70; the index of dispersion comes to $219.\dot{8}\dot{1}$. If we ignore the recommendation that χ^2 should only be relied on if the expectation in each cell exceeds 5, and apply it to a case in which the expectation is only 0.157, we have

$$\begin{array}{rr} \sqrt{2\chi^2} & 20.9675 \\ \sqrt{2n-1} = \sqrt{277} & \underline{16.6433} \\ \text{Difference} & 4.3242 \end{array}$$

The difference indicates strong significance, at a level of about one in a hundred thousand ($P < .00001$), but the experimenter is still left in some doubt as to its validity.

The conventional application of the more comprehensive χ^2 test is at least equally embarrassing. Using the sufficient estimate, $m = \bar{r}$, we have

	observed	expected	ungrouped	grouped
r	a	m_r	$(a - m_r)^2/m_r$	$(a - m_r)^2/m_r$
0	124	119.6415	.1588	.1588
1	12	18.8008	2.4600	2.4600
2	2	1.4772	.1850	
3	2	.0774	47.76	3.8293
>3	0	.0031	.0031	
Total	140	140.0000	50.57 $n = 3$	6.4481 $n = 1$

The χ^2 test with three degrees of freedom suggests high significance, owing to the two observations of the variate 3, but the expectation in this class is obviously too low for the test to be valid. The test with one degree of freedom shows significance just beyond the 1% point.

Some, however, will object that the expectation in the third class is still only 1.5577, while others may feel that the test has been made unduly insensitive by pooling the twos with the threes, thereby ignoring the strong evidence which the latter really provide.

THE EXACT FREQUENCY DISTRIBUTION

In a rigorous treatment of the problem we must distinguish between two different facets of it. (i) The calculation of an exact frequency distribution giving the objective frequencies of the observational series we have obtained, and of other alternative observational series. (ii) The choice of a criterion by which the assemblage of possibilities is to be divided up in making a test of significance. The first facet is purely mathematical; only one answer is possible. The second takes account of the needs and prior ideas of the experimenter; by it he specifies the particular question he chooses to ask. The two procedures illustrated in Section 2 attempted answers to two different questions: (a) Is the sample more variable than a sample with this mean should be, and if so at what level of significance; (b) Do the frequencies observed differ significantly from those expected in a Poisson Series, having the same mean, and if so at what level of significance. With an exact distribution other questions may suggest themselves.

The key to the distributional problem is that the only parameter of the Poisson Series is capable of sufficient estimation. From this it follows that the Likelihood function of m is identical for all observational series having the same size and the same mean $\bar{r}$. The relative frequencies of all such series are therefore entirely independent of the unknown parameter. These frequencies can be calculated. There will, it is true, be a great many of them equal indeed to the number of partitions of the partible number 22, but we shall not need to calculate all of these, especially of the rarer, in order to answer such questions as: How frequently is

$$S\{a_r(r - \bar{r})^2\}/\bar{r}$$

less than $219.\dot{8}\dot{1}$, or, more expeditiously, how often is

$$S_r(a_r r^2)$$

less than 38. Or, if we consider the second test, to answer the question: How frequently is

$$S(a_r - m_r)^2/m_r$$

less than 50.57. If either of these are the questions of our choice, they can clearly be answered.

If, for any value m of the unknown parameter

$$p_r = e^{-m} m^r / r!,$$

then the probability of an observational sequence a_0 , a_1 , a_2 , $\cdots$ is

$$\frac{A!}{a_0!\, a_1!\, a_2! \cdots} p_0^{a_0} p_1^{a_1} p_2^{a_2} \cdots$$

Whence, substituting for p_r , we have the constant function of m

$$e^{-Am} m^B$$

with the particular factor of the distribution, which is independent of m,

$$\frac{A!}{a_0!\, a_1!\, a_2! \cdots} \cdot \frac{1}{(2!)^{a_2} (3!)^{a_3} \cdots} .$$

Moreover the sum of all such probabilities will be the probability of observing the number B in a Poisson series with a parameter Am, and is

$$e^{-Am} \frac{(Am)^B}{B!} .$$

Consequently, the relative frequency of any observable series, having A and B as specified, will be

$$\frac{B!}{A^B} \cdot \frac{A!}{a_0!\, a_1!\, a_2! \cdots} \cdot \frac{1}{(2!)^{a_2} (3!)^{a_3} \cdots} ,$$

and this we can calculate without knowledge of m.

A LIST OF CONFIGURATIONS

The possible observed distributions having $A = 140$, $B = 22$ correspond to the partitions of the partible number 22. If A were less than B, we should also stipulate that the number of parts should not exceed A. The correspondence may be shown in a short table

Partition	a_0	a_1	a_2	a_3
(1^{22})	118	22	—	—
(21^{20})	119	20	1	—
$(2^2 1^{18})$	120	18	2	—
(31^{19})	120	19	—	1

and so on. Corresponding with any distribution we can now list the corresponding frequencies, and any criterion value the use of which may be in view, as in Table 1.

The list of the first forty-five partitions in order of frequency extends, further than is needed for any ordinary criteria. The reason for this is that using a variety of tests, different criteria are liable to put the distributions in different orders, so that when treated exactly, slightly different questions receive slightly different answers.

The first point to be noted is that the probability of observing exactly the series we have observed is .000222; consequently on no view is the level of significance higher than one in 4500. It should not, however, be placed so high as this, since on any reasonable view some other possible series will be judged more discrepant than the one we have observed.

Next, the sum of the probabilities of the series observed together with all less frequent series is .000639. This gives a possible test of significance at a level of about one in 1600. It is not a good test for general use, but seems to be not inappropriate for the present problem. The kind of case in which it fails to carry conviction is the not uncommon one in which it is a matter of arbitrary choice whether a class of observational series, which may be distinguished, but which seem to differ in nothing relevant to the actual problem under discussion, shall be considered separately or together; for this choice will decide its order relative to other classes of observation. Mere order of frequency cannot therefore in general be taken to provide an order of discrepancy.

The standard large-sample χ^2 tests obviously provide valid, but, for our purpose, somewhat conventional tests. The index of dispersion depends on the integer $S(r^2a_r)$, which for our sample is 38. The probability of observing a value 38 or higher is .001112, at a level of about one in 900. But the probability of observing actually 38 is .0007355, and the fact that this is a large fraction of the total, while not invalidating the test of significance, shows that it is somewhat less sensitive than it might be, for a reason extraneous to our purpose, namely, that it is somewhat heavily grouped. Of course, if our purpose is sharply defined as "to test if the sample variance is too high for a Poisson series, and if so at what level of significance," this limitation is inherent in the problem we have set ourselves.

For the general χ^2 test the probability of χ^2 as great as or greater than 50.499 is .002684, or rather more than one in 400. In comparison with the previous result this shows that though on either test the significance of the deviation is high, yet the deviations are particularly

TABLE 1

Partition	Frequency	$S(r^2 a_r)$	χ^2	$S(a \log (ar!))$ -273.222
(21^{20})	.340167	24	0.315	0.090
$(2^2 1^{18})$	.269299	26	0.301	0.079
(1^{22})	.175238	22	2.125	0.793
$(2^3 1^{16})$	.113506	28	2.083	0.295
$(2^4 1^{14})$	.027911	30	5.661	0.972
(321^{17})	.026707	30	11.346	0.792
(31^{19})	.018898	28	12.485	1.354
$(32^2 1^{15})$	.014886	32	12.004	0.938
$(32^3 1^{13})$	.004236	34	14.458	1.430
$(2^5 1^{12})$	.004130	32	11.036	1.787
(421^{16})	.000930	36	324.66	2.261
(41^{18})	.000742	34	328.57	2.770
$(32^4 1^{11})$	.000666	36	18.708	2.208
$(3^2 1^{16})$	.000620	34	49.716	2.738
$(3^2 21^{14})$	.000605	36	49.249	2.342
$(42^2 1^{14})$	.000454	38	325.84	2.466
$(2^6 1^{10})$	.000366	34	18.207	2.837
$(3^2 2^2 1^{12})$	.000222	38	50.499	2.675
$(42^3 1^{12})$	.000111	40	331.23	3.027
$(32^5 1^9)$	.000059	38	24.754	3.258
	per million			
(431^{15})	40.34	40	340.31	3.486
$(3^2 2^3 1^{10})$	39.08	40	53.705	3.385
(4321^{13})	34.16	42	340.07	3.303
(521^{15})	24.29	44	10466.	3.958
(51^{17})	21.97	42	10467.	4.311

of a kind to increase the variance, rather than in other ways to increase the general value of χ^2. The contrast is heightened by the fact that the general χ^2 is not grouped.

It will be seen, however, that the conventional χ^2 value is absurdly sensitive to high observations such as 4 and 5, and that such occurrences govern its value. A better general test when expectations are liable to be very small is provided by

$$\underset{r}{S}\, a_r \log \frac{a_r}{m_r},$$

which, if the logarithms are to the base e, agrees with $\frac{1}{2}\chi^2$

TABLE 1—*Continued*

Partition	Frequency	$S(r^2a_r)$	χ^2	$S(a \log (ar!))$ -273.222
$(2^7 1^8)$	18.84	36	27.173	4.139
$(42^4 1^{10})$	14.66	42		
$(432^2 1^{11})$	10.66	44		
$(52^2 1^{13})$	10.29	46		
$(3^3 1^{13})$	7.59	40		
$(3^3 21^{11})$	4.74	42		
$(3^2 2^4 1^8)$	3.49	42		
$(32^6 1^7)$	2.79	40	32.597	
$(52^3 1^{11})$	2.14	48		
$(432^3 1^9)$	1.55	46		
$(42^5 1^8)$	1.05	44		
$(3^3 2^2 1^2)$	1.03	44		
(531^{14})	.98	48		
(5321^{12})	.71	50		
$(4^2 1^{14})$	.61	46		
$(43^2 1^{12})$	.59	46		
$(2^8 1^6)$	.50	38	37.936	
(61^{16})	.50	52		
(621^{14})	.49	54		
$(4^2 21^{12})$	.44	48		

when the values of m are large enough for the latter to be valid.

For convenience I have taken logarithms to the base 10, and use in the last column the easily calculable criterion

$$\underset{r}{S} \, (a_r \log_{10} (a_r r!)) - 273.222;$$

my values would therefore have to be multiplied by 2 $\log_e$ 10 or 4.6 to give values comparable to χ^2.

This measure of discrepancy also is not grouped. The value for our sample is 2.675 and the probability of a value as great as this or greater is .002367, at a rather higher level of significance than the large sample χ^2 test when treated exactly.

SUMMARY

These tests make it clear that the sample departs from expectation based on the Poisson Series at a level, for a general test, of about $\frac{1}{4}\%$;

while a corresponding refinement of the Index of Dispersion, the measure of variance relative to the mean, claims a level of about one in 900, in spite of the fact that this measure, owing to the grouping of different samples, is inherently somewhat insensitive. It is in the direction of increased dispersion that our sample departs most distinctly from the Poisson expectation.

The use is illustrated of a generalised measure of deviation,

$$S\left\{a \log\left(\frac{a}{m}\right)\right\},$$

which is essentially the logarithmic difference in likelihood between the most likely Poisson Series and the most likely theoretical series without restriction. This measure seems well fitted to take the place of the conventional χ^2, when class expectations are small.

FIGURE 1

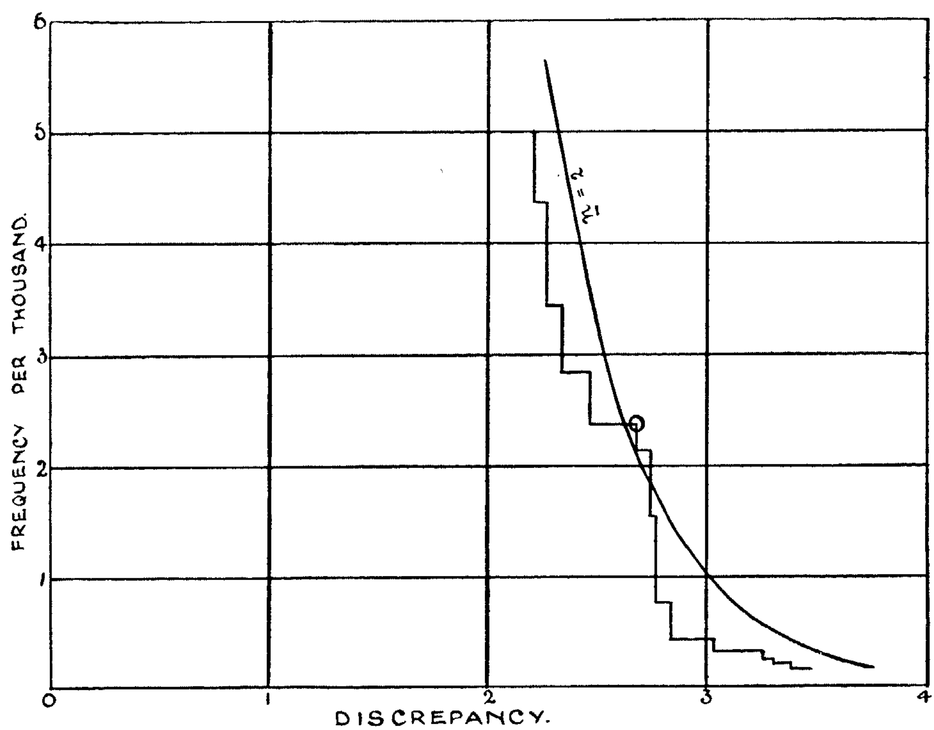

Figure 1 compares the actual discontinuous distribution of this criterion with that of the theoretical χ^2 for two degrees of freedom.