

For details of other non-parametric methods, and in particular of the Mann-Whitney  $U$  test, which is in many ways more satisfactory than the tests based on medians, the reader is referred to Siegel's book *Non-parametric Statistics*. These methods have acquired a considerable vogue, particularly among social scientists and psychologists, but in my view they are rarely necessary because of the robustness of most parametric tests, and they have the serious disadvantage of being less flexible than parametric methods, so that it is difficult to adapt them to suit the situation and it is difficult to derive confidence intervals from them.

THE  $\chi^2$  TEST OF GOODNESS OF FIT

A problem which frequently arises is that of testing the agreement between observation and hypothesis. The most useful measure of agreement, which can be applied whenever the observations can be grouped either naturally or artificially into a finite number of classes, is the  $\chi^2$  criterion devised by Karl Pearson (1900). Suppose then that we have made  $n$  observations which can be grouped into  $k$  classes. We will write  $n_i$  for the observed number of observations and  $E_i$  for the Expected number of observations in the  $i$ th class; thus  $E_i = nP_i$  where  $P_i$  is the probability, calculated from our hypothesis, that an observation will fall in the  $i$ th class. The  $\chi^2$  criterion of goodness of fit is defined as

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - E_i)^2}{E_i}$$

If there is perfect agreement, then  $n_i = E_i$  for all  $i$  and  $\chi^2 = 0$ ; the worse the fit, the larger  $\chi^2$  will be.

For example, we find from Table 11 on p. 82 that the observed and Expected frequencies of the number of heads in 2000 sequences of 5 spins of a coin are:

No. of heads	0	1	2	3	4	5	Total
Observed frequency	59	316	596	633	320	76	2000
Expected frequency	62.5	312.5	625	625	312.5	62.5	2000
Difference	-3.5	+3.5	-29	+8	+7.5	+13.5	0

For these data

$$\chi^2 = \frac{3.5^2}{62.5} + \frac{3.5^2}{312.5} + \dots + \frac{13.5^2}{62.5} = 4.78.$$

In this example the theoretical probabilities,  $P_i$ , are known exactly. In most applications, however, these probabilities involve one or more unknown parameters which must be estimated from the observations before the Expected numbers can be calculated. Thus in Table 12 on p. 89 the probability of a male birth had to be estimated from the observed proportion of boys before the Expected numbers of families with different numbers of boys in them could be calculated from the binomial formula; similarly in Table 17 on p. 113 the mean and variance of the observed frequency distribution had to be calculated before a normal distribution could be fitted. The fact that one or more parameters have been estimated from the data before the Expected numbers could be obtained does not affect the way in which the  $\chi^2$  criterion is calculated; for example, for the data in Table 12

$$\chi^2 = \frac{(215 - 165)^2}{165} + \dots + \frac{(342 - 264)^2}{264} = 92.1.$$

It does, however, affect the sampling distribution of the criterion and in consequence the interpretation to be placed upon it once it has been calculated.

An important group of applications of the  $\chi^2$  criterion is in testing for independence in contingency tables. On p. 13 we considered the following data on the sex and viability of births in England and Wales:

	Liveborn	Stillborn	Total
Male	359,881 (360,056)	8,609 (8,434)	368,490
Female	340,454 (340,279)	7,796 (7,971)	348,250
Total	700,335	16,405	716,740

Such a table is called a  $2 \times 2$  contingency table since each of the characters (sex and viability) is divided into two classes. If sex and viability were independent of each other the probability of a male livebirth would be the product of the overall

probabilities of these two events, which can be estimated from the corresponding marginal proportions. Hence the Expected number of male livebirths, supposing these factors to be independent, is calculated as

$$716740 \times \frac{368490}{716740} \times \frac{700335}{716740} = 360,056.$$

The other Expected numbers can be calculated in a similar way and are shown in brackets. The difference between observed and Expected numbers is in each case 175 in absolute value, so that the  $\chi^2$  criterion for departure from the hypothesis of independence is

$$\chi^2 = \frac{175^2}{360056} + \frac{175^2}{340279} + \frac{175^2}{8434} + \frac{175^2}{7971} = 7.56.$$

We have seen that if there is perfect agreement between observation and hypothesis  $\chi^2 = 0$  and that the worse the agreement the larger  $\chi^2$  is. In order to interpret an observed value of this criterion we must know its sampling distribution on the assumption that the hypothesis being tested is true. We shall now show that the  $\chi^2$  criterion follows *approximately* the  $\chi^2$  distribution with  $k-1-\beta$  degrees of freedom, where  $k$  is the number of classes into which the observations are divided and  $\beta$  is the number of parameters which have been *independently* estimated from the data; it is from this fact that the  $\chi^2$  distribution gets its name. We must imagine that a large number of experiments have been performed in each of which  $n$  observations, classified into  $k$  groups, have been obtained and the  $\chi^2$  criterion calculated; what is the probability distribution generated by the different values of  $\chi^2$  assuming that the hypothesis being tested is true? We consider first the case in which the probabilities,  $P_i$ , are specified completely by hypothesis and no parameters need to be estimated.

Suppose then that  $n$  observations have been made and that the probability that a particular observation will fall into the  $i$ th class is  $P_i$ . The number of observations falling in the  $i$ th class will vary from experiment to experiment and is therefore a random variable. By an extension of the argument used in deriving the binomial distribution it is quite easy to

show that the probability that  $n_1$  observations will fall into the first class,  $n_2$  into the second class and so on, is

$$P(n_1, n_2, \dots, n_k) = \frac{n!}{n_1!n_2!\dots n_k!} P_1^{n_1} P_2^{n_2} \dots P_k^{n_k}$$

where, of course

$$P_1 + P_2 + \dots + P_k = 1 \\ n_1 + n_2 + \dots + n_k = n.$$

This distribution is called the multinomial distribution (see Problem 6.10). The binomial distribution is a special case of the multinomial distribution with  $k = 2$ .

The numbers of observations are not independent random variables but are negatively correlated; for on the occasions when the number of observations in the first class, for example, is above average we should expect the other numbers to be on the average rather low since their sum is fixed. In the case of the binomial distribution there is complete negative correlation since  $n_2 = n - n_1$ . The  $n_i$ 's can, however, be regarded as independent Poisson variates with  $\mu_i = nP_i$  subject to the restriction that their sum is  $n$ ; that is to say, their distribution is the same as that which would be obtained if we took  $k$  independent Poisson variates with these means and then discarded all the occasions on which they did not add up to the fixed number  $n$ . For the joint probability distribution of these independent Poisson variates is

$$\frac{e^{-nP_1} (nP_1)^{n_1}}{n_1!} \dots \frac{e^{-nP_k} (nP_k)^{n_k}}{n_k!} = e^{-n \sum P_i} \frac{P_1^{n_1} \dots P_k^{n_k}}{n_1! \dots n_k!}$$

and the probability that they will add up to  $n$  is

$$\frac{e^{-n \sum P_i}}{n!}$$

since, by the additive property of Poisson variates,  $\sum n_i$  is itself a Poisson variate with mean  $\sum nP_i = n$ . If we divide the first expression by the second to obtain the conditional probability distribution of the  $n_i$ 's given that they add up to  $n$ , we get the multinomial distribution.

Suppose then that  $n_1, n_2, \dots, n_k$  are independent Poisson variates with means  $nP_1, nP_2, \dots, nP_k$ ; then

$$\zeta_i = \frac{(n_i - nP_i)}{\sqrt{nP_i}}$$

has zero mean and unit variance and is approximately normally distributed provided that  $nP_i$  is not too small. Let us now make an orthogonal transformation from the  $\zeta_i$ 's to a new set of variables  $\mathcal{Y}_i$  in which

$$\mathcal{Y}_1 = \sum_{i=1}^k \sqrt{P_i} \zeta_i$$

and let us impose the restraint that  $\sum_{i=1}^k n_i = n$  which is equivalent to  $\mathcal{Y}_1 = 0$ . It follows from the theorem in the last chapter that

$$\chi^2 = \sum_{i=1}^k \zeta_i^2 = \sum_{i=2}^k \mathcal{Y}_i^2$$

follows approximately the  $\chi^2$  distribution with  $k-1$  degrees of freedom. For this approximation to hold it is necessary that the Expected values,  $nP_i$ , should not be too small; it has been found empirically that the approximation is satisfactory, provided that each of them is greater than 5.

So far we have considered only the case when the probabilities are specified completely by hypothesis. If one or more parameters have to be estimated from the data, this will clearly decrease the average value of  $\chi^2$  since they will be estimated to make the fit as good as possible. It can in fact be shown that, provided the parameters are estimated in a reasonable way, each independent parameter estimated is equivalent to placing an additional linear restraint on the observations. Hence, if  $p$  parameters are independently estimated,  $\chi^2$  will follow approximately the  $\chi^2$  distribution with  $k-1-p$  degrees of freedom.

It should be noted that in the  $2 \times 2$  contingency table considered on p. 155 only two parameters, and not four, have been independently estimated from the data since, once the probability of a male birth has been estimated as  $368490/716740 = .5141$ , it follows immediately that the probability of a female birth

will be estimated as  $1 - .5141 = .4859$  and likewise for live-birth *v.* stillbirth; the  $\chi^2$  criterion for testing the independence of these two factors therefore follows a  $\chi^2$  distribution with  $4-2-1 = 1$  degree of freedom. In the general case of an  $r \times s$  contingency table in which the first character is divided into  $r$  classes and the second into  $s$  classes only  $r+s-2$  marginal probabilities are independently estimated since the last relative frequency in the row margin is known once the previous  $r-1$  relative frequencies have been calculated, and likewise for the column margin. Hence the  $\chi^2$  criterion of independence follows the  $\chi^2$  distribution with  $rs-1-(r+s-2) = (r-1)(s-1)$  degrees of freedom.

Several experiments have been done to verify the foregoing theory. In one set of experiments Yule threw 200 beans into a revolving circular tray with 16 equal radial compartments and counted the number of beans falling into each compartment. The 16 frequencies so obtained were arranged (1) in a  $4 \times 4$  table, and (2) in a  $2 \times 8$  table. Then  $\chi^2$  was calculated as for testing independence in a contingency table. This experiment was repeated 100 times; the observed and theoretical distributions are shown in Table 20.

TABLE 20  
Distribution of  $\chi^2$  in Yule's experiment with beans  
(Yule and Kendall, 1950)

$\chi^2$	$4 \times 4$		$8 \times 2$	
	Observed	Expected (9 d.f.)	Observed	Expected (7 d.f.)
0-5	17	17	30	34
5-10	44	48	56	47
10-15	32	26	10	15
15-20	6	7	3	3
Over 20	1	2	1	1
Total	100	100	100	100

A less artificial realisation of the distribution of the  $\chi^2$  criterion may be found in a paper by Chamberlain and Turner (1952). These authors did duplicate white cell counts on 294