APPLICATIONS OF NEYMAN'S $C(\alpha)$ TECHNIQUE

F. N. DAVID
UNIVERSITY OF CALIFORNIA, RIVERSIDE

1. Introduction

The χ^2 techniques are possibly the most widely used in statistical methods in that they are simple in application and interpretation. This very generality, however, leads to a lack of sensitivity of the test criterion since the hypotheses alternate to that under test are commonly only vaguely specified. During the past few years there have been procedures put forward which enable more sensitive tests to be made. We give here a series of models for the alternative hypothesis under χ^2 type situations and the appropriate test criteria. The technique used to derive the test criteria is that put forward by Neyman [1] and loosely referred to by us as the $C(\alpha)$ procedure. The power functions of these tests may be calculated, and in such cases as we have investigated, they show the chosen criteria to be comparable in sensitivity to others which may be proposed.

2. Models reflecting a change in the location parameter

Suppose two groups of individuals with n in the first group (A) and N in the second group (B). On each individual the same characteristic X is measured. These measurements are used to divide the groups into $s + 1$ categories, so that we have Table I.

TABLE I

	Categories of measurement				
	1	2	$\cdots$	$s+1$	Total
A	n_1	n_2	$\cdots$	n_{s+1}	n
B	N_1	N_2	$\cdots$	N_{s+1}	N
Totals	M_1	M_2	$\cdots$	M_{s+1}	M

It is apparent that if the measurements of the characteristic X are accurate and there are no "tied" values, a rank sum criterion is the appropriate quantity

This investigation was partially supported by USPHS Research Grant No. GM-10525-08, National Institutes of Health, Public Health Service.

to test for possible differences in the location parameters of the populations generating groups A and B. There are, however, many situations in experimental data where ties are abundant, particularly in the first (or last) categories where often all that is recorded is that the measurement was less (or greater) than a given value. The χ^2 criterion will test the null hypothesis that there is no difference between the population location parameters, but it will be sensitive to all alternative hypotheses instead of just that of a possible difference. A model set up by J. Neyman [2] would appear to be appropriate for this situation.

Let the hypothesis alternative to the null be that the location parameter of the population generating group B is greater than that of the population generating group A. Accordingly, we postulate the setup in Table II with $\xi = 0$ for the null hypothesis, and the obvious restraints on ξ for the alternate hypothesis.

TABLE II

	Categories of measurement 1	2	3	$\cdots$	s	$s+1$	Total
A	p_1	p_2	p_3	$\cdots$	p_s	p_{s+1}	1
B	$p_1(1-\xi)$	$p_2 + \xi(p_1 - p_2)$	$p_3 + \xi(p_2 - p_3)$	$\cdots$	$p_s + \xi(p_{s-1} - p_s)$	$p_{s+1} + \xi p_s$	1

Using Neyman's procedure we have, conditional on n and N, that the test criterion T is

$$(1) \qquad T = \sum_{j=1}^{s} (p_{j-1} - p_j)\left[\frac{N_j}{p_j} - \frac{N_{s+1}}{p_{s+1}}\right] - \frac{N}{M}\sum_{j=1}^{s} (p_{j-1} - p_j)\left[\frac{M_j}{p_j} - \frac{M_{s+1}}{p_{s+1}}\right]$$

with $p_0 = 0$. Writing

$$(2) \qquad A_r = \sum_j \frac{(p_{j-1} - p_j)^{r+1}}{p_j^r} + \frac{p_s^{r+1}}{p_{s+1}^r},$$

we have, under H_0,

$$(3) \qquad \begin{aligned} &\mu_1'(T) = 0, && \mu_2(T) = \frac{nN}{M} A_1, \\ &\mu_3(T) = \frac{nN}{M}\left(\frac{n^2 - N^2}{M^2}\right) A_2, && \kappa_4(T) = \frac{nN}{M}\left(\frac{n^3 + N^3}{M^3}\right)[A_3 - 3A_1^2], \end{aligned}$$

indicating that for reasonable sized M the approximation to normality is satisfactory.

3. Power function of the test

Conditional on n and N as before, we have under H_1 ($\xi > 0$) that

$$(4) \qquad \mu_1'(T) = \frac{nN}{M}[A_1\xi]$$

and

$$(5) \qquad \mu_2(T) = \frac{nN}{M}\left[A_1 + \xi A_2 \frac{n}{M} - \xi^2 A_1^2 \frac{n}{M}\right].$$

Consequently, following the usual procedure and assuming that T is normally distributed, the power of the test may be computed.

4. Approximations involved in the test

The criterion T may be written as

$$(6)\qquad T = \frac{1}{M}\left[\sum_{j=1}^{s} (nN_j - Nn_j)\left(\frac{p_{j-1} - p_j}{p_j}\right) + (nN_{s+1} - Nn_{s+1})\frac{p_s}{p_{s+1}}\right]$$

with $p_0 = 0$. The parameters p will not generally be known and estimates of them will usually be substituted. Since $\xi = 0$ under the null hypothesis, we put $p_j = M_j/M$ obtaining

$$(7)\qquad \hat{T} = \frac{1}{M}\left[\sum_{j=1}^{s} (nN_j - Nn_j)\left(\frac{M_{j-1} - M_j}{M_j}\right) + (nN_{s+1} - Nn_{s+1})\frac{M_s}{M_{s+1}}\right].$$

It is easy to show that $\mathcal{E}\hat{T} = 0$ under the null hypothesis. Further, if we write for the denominator

$$(8)\qquad \frac{nN}{M^2}\left(\sum \frac{(M_{j-1} - M_j)^2}{M_j} + \frac{M_s^2}{M_{s+1}}\right),$$

it has expected value

$$(9)\qquad \frac{nN}{M}\left[\left(\sum_{j=1}^{s} \frac{(p_{j-1} - p_j)^2}{p_j} + \frac{p_s^2}{p_{s+1}}\right) + \frac{1}{M}\sum_{j=1}^{s+1}\left(\frac{p_{j-1}^2}{p_j^2} + \frac{p_{j-1}}{p_j}\right)\right]$$

if we exclude terms of higher order than $1/M$. This indicates that substituting the estimates of the p will tend to make the denominator too large, on the average, and the nominal significance levels will, accordingly, underestimate significance, a fault on the right side. To order $1/M$ the covariance between the numerator and the denominator is zero.

5. Neyman's randomization approach

Neyman [2] put forward the model given above but with an essential difference. He assumed that a total of M elements was given, but that the dichotomy of the M into the two samples A and B was achieved as the result of a random experiment. Once the dichotomy is made the n and N elements were each subjected to measurement and categorization as described above. The test criterion

$$(10)\qquad T_R = \sum_{j=1}^{s} (p_{j-1} - p_j)\left(\frac{N_j}{p_j} - \frac{N_{s+1}}{p_{s+1}}\right) - \Pi \sum_{j=1}^{s} (p_{j-1} - p_j)\left(\frac{M_j}{p_j} - \frac{M_{s+1}}{p_{s+1}}\right),$$

where Π is the probability that one of the M elements is assigned to the sample, is the same as in the conditional case if we put $\Pi = N/M$. The variance of T_R is

$$(11)\qquad \sigma_{T_R}^2 = \Pi(1 - \Pi)\sum_{j=1}^{s} \frac{(p_{j-1} - p_j)^2}{p_j} + \frac{p_s^2}{p_{s+1}}$$

and as before, the assumption is made that (T_R/σ_{T_R}) is a unit normal variable. The expectation of T_R is zero under H_0. Clearly, Π is determined by the initial random experiment.

Under the alternate hypothesis

$$\mathcal{E}(T_R|\xi) = \xi\Pi(1 - \Pi)\left(\sum_{j=1}^{s} \frac{(p_{j-1} - p_j)^2}{p_j} + \frac{p_s^2}{p_{s+1}}\right) \tag{12}$$

and

$$\operatorname{Var}(T_R|\xi) = \operatorname{Var}(T_R|\xi = 0) + \xi\Pi(1 - \Pi)^2\left[\sum_{j=1}^{s} \frac{(p_{j-1} - p_j)^3}{p_j^2} + \frac{p_s^3}{p_{s+1}^2}\right] - \xi^2[\operatorname{Var}(T_R|\xi = 0)]^2. \tag{13}$$

6. Example (Neyman's random experiment)

In an experiment to determine whether the larvae of *Lema triliveata daturaphila* (a kind of potato beetle) have a sense of smell, a target was made with an attractant at the "bull's eye." Twenty larvae were released individually 13 mm from the target center and the performance of each was classified as being nondirectional, intermediate, and directional, the dichotomy into classes being made on the basis of the distance traveled and the time taken. Twenty more larvae were released individually from a target distance of 23 mm and the same classification of performance made. The results of the experiments were as shown in Table III.

TABLE III

PERCEPTION TESTS FOR LARVAE

Distance from target center (mm)	Categories Nondirectional	Intermediate	Directional	Total
13	2 (N_1)	5 (N_2)	13 (N_3)	20
23	4 (n_1)	7 (n_2)	9 (n_3)	20
Totals	6 (M_1)	12 (M_2)	22 (M_3)	40

Under the null hypothesis of no difference of sense of smell with distance there should be no real difference between the performances at the two distances. Under the alternate hypothesis the larvae should perform better at 13 mm than they do at 23 mm. Here we have $s = 2$, $N = n = 20$, and the test criterion is

$$-(N_1 - \Pi M_1) - (N_2 - \Pi M_2)\left(1 - \frac{p_1}{p_2}\right) + \frac{p_2}{p_3}(N_3 - \Pi M_3) \tag{14}$$

or, taking $\Pi = \frac{1}{2}$ and substituting estimates for the p,

$$\frac{1}{2}\left[-(N_1 - n_1) + \left(\frac{M_1 - M_2}{M_2}\right)(N_2 - n_2) + \frac{M_2}{M_3}(N_3 - n_3)\right]. \tag{15}$$

Calculations give an equivalent normal deviate of less than unity indicating that there is no significance.

7. Example (conditional test)

Fisher [3] gives data concerning the infestation of sheep by ticks for two different distributions. We adapt his data by dividing the distributions into four categories, namely, light infestation, medium, fairly heavy, and very heavy, as given in Table IV. The conditional criterion based on the $C(\alpha)$ technique will be appropriate here. Either the $C(\alpha)$ criterion or the χ^2 criterion is very significant indicating that the population of which II is a sample is probably more heavily infested than that of I.

TABLE IV

Sets of sheep	Infestation Light	Medium	Fairly heavy	Very heavy	Total
I	24	21	12	3	60
II	20	19	19	24	82
Totals	44	40	31	27	142

8. Differences in dispersion parameters

It is clear from the setup of Table II that the alternative to the hypothesis under test is equivalent to that of a difference in the location parameters of the two populations generating the samples. This indicates the possibility of writing variants on the setup which will allow the alternative to the hypothesis under test to be equivalent to a difference in the dispersion parameters of the two populations. Such a situation frequently arises in experimental work when two treatments designed to have the same average effect are being compared and it is desired to know whether one treatment is more variable than the other.

Under the alternate hypothesis there will be (possibly) a diminution of probabilities at the tails when one population is compared with the other, and a heaping up at the center. Assume that there is an even number of categories. (This is not an important restriction and the result for an odd number of categories can easily be obtained.) We write, for $\xi > 0$, the setup in Table V.

TABLE V

Population	Categories 1	2	$\cdots$	s	$s+1$	$\cdots$	$2s$	Total
A	p_1	p_2	$\cdots$	p_s	p_{s+1}	$\cdots$	p_{2s}	1
B	$p_1 - \xi p_1$	$p_2 + \xi(p_1 - p_2)$	$\cdots$	$p_s + \xi p_{s-1}$	$p_{s+1} + \xi p_{s+2}$	$\cdots$	$p_{2s}(1 - \xi)$	1

If the sample values are

$$(16)\qquad \sum_{j=1}^{2s} n_j = n, \qquad \sum_{j=1}^{2s} N_j = N, \qquad n + N = M,$$

we have conditionally that, assuming $M_0 = 0 = M_{2s+1}$,

$$(17)\quad T = \sum_{j=1}^{s-1} \left(\frac{M_{j-1} - M_j}{M_j}\right)\left[N_j - \frac{NM_j}{M}\right] + \frac{M_{s-1}}{M_s}\left(N_s - \frac{N}{M} M_s\right) + \frac{M_{s+2}}{M_{s+1}}\left(N_{s+1} - \frac{N}{M} M_{s+1}\right) + \sum_{t=s+2}^{2s} \frac{(M_{t+1} - M_t)}{M_t}\left(N_t - \frac{N}{M} M_t\right)$$

and

$$(18)\quad \operatorname{Var} T = \frac{Nn}{M}\left(\sum_{j=1}^{s} \frac{(M_{j-1} - M_j)^2}{M_j} + \frac{M_{s-1}^2}{M_s} + \frac{M_{s+2}^2}{M_{s+1}} + \sum_{t=s+2}^{2s} \frac{(M_{t+1} - M_t)^2}{M_t}\right).$$

Since the expected value of T is zero under H_0, the test criterion is $T/(\operatorname{Var} T)^{1/2}$, significance being judged from normal tables.

Under H_1 we have

$$(19)\quad T = \frac{nN\xi}{M}\left[\sum_{j=1}^{s-1} \frac{(p_{j-1} - p_j)^2}{p_j} + \sum_{t=s+2}^{2s} \frac{(p_{t+1} - p_t)^2}{p_t} + \frac{p_{s-1}^2}{p_s} + \frac{p_{s+2}^2}{p_{s+1}}\right]$$

and

$$(20)\quad \operatorname{Var} T = \frac{nN}{M}\left[\sum_{j=1}^{s-1} \frac{(p_{j-1} - p_j)^2}{p_j} + \sum_{t=s+2}^{2s} \frac{(p_{t+1} - p_t)^2}{p_t} + \frac{p_{s-1}^2}{p_s} + \frac{p_{s+2}^2}{p_{s+1}}\right] + \frac{n^2N}{M^2}\xi\left[\sum_{j=1}^{s-1} \frac{(p_{j-1} - p_j)^3}{p_j^2} + \sum_{t=s+2}^{2s} \frac{(p_{t+1} - p_t)^3}{p_t^2} + \frac{p_{s-1}^3}{p_s^2} + \frac{p_{s+2}^3}{p_{s+1}^2}\right] - N[\mathcal{E}(T)]^2,$$

so that the power of the test to detect a positive value of ξ may be computed. It will be recognized that the choice of a central dichotomy—in this case I have made a dichotomy between the sth and the $(s + 1)$st categories—is an arbitrary one. Further, if there is a change in the location parameter at the same time as the change in the dispersion parameter, the first change will tend to mask the second. This is, however, a situation which is always present in "nonparametric" tests for dispersion.

9. Number of groups

Commonly in the use of χ^2 and allied tests, the number of groups used for calculating the test criterion is fixed by the conditions under which the data are gathered. It is, therefore, useful to look at such $C(\alpha)$ tests as may be regarded as competitors of χ^2 with the purpose of determining the number of groups which will give optimum power. Calculations were carried out regarding the power, assuming equal probabilities for each group under H_0, and letting $n = N = M/2$. Under these assumptions, the conclusions reached were that the smallest number of groups consistent with the H_1 model should be aimed for. Thus, we have the following.

(i) *Change in location.*
Models:

$$H_0 \quad \begin{array}{cc|c} p_1 & 1-p_1 & 1 \\ p_1 & 1-p_1 & 1 \end{array} \qquad H_1 \quad \begin{array}{cc|c} p_1(1-\xi) & 1-p_1(1-\xi) & 1 \\ p_1 & 1-p_1 & 1 \end{array}$$

Sample setup:

N_1	N_2	N
n_1	n_2	n
M_1	M_2	M

The test criterion is

(21) $$T = \frac{Nn_1 - nN_1}{M_2} = \frac{M_1N_2 - N_1M_2}{M_2}$$

and

(22) $$\operatorname{Var} T = \frac{nN}{M}\frac{M_1}{M_2}.$$

(ii) *Change in dispersion.*
Models:

$$H_0 \quad \begin{array}{ccc|c} p_1 & p_2 & p_3 & 1 \\ p_1 & p_2 & p_3 & 1 \end{array} \qquad H_1 \quad \begin{array}{ccc|c} p_1 & p_2 & p_3 & 1 \\ p_1(1-\xi) & p_2 + \xi(p_1 + p_2) & p_3(1-\xi) & 1 \end{array}$$

Sample setup:

n_1	n_2	n_3	n
N_1	N_2	N_3	N
M_1	M_2	M_3	M

The test criterion is

(23) $$T = \frac{nN_2 - Nn_2}{M_2}$$

and

(24) $$\operatorname{Var} T = \frac{nN(M_1 + M_3)}{MM_2}.$$

10. Model for fatal road accidents

It is clear that the alternative hypothesis H_1 can be built suitable to the conditions of any particular problem provided that this building is done before any set of data is scrutinized. For example, suppose we remember that the age and sex of pedestrians killed by moving vehicles are both recorded and published. It may be desired to test whether the age distribution for males is the same as the age distribution for females. An opinion is expressed that more little boys would tend to be killed than little girls because they are more active and, therefore, have more exposure to risk and similarly that more of the aged killed would be men. The Ministry of Transport's (U.K.) Report on Fatal Road Accidents (1922) gives the information in Table VI.

TABLE VI

AGE AND SEX OF PEDESTRIANS KILLED

Sex	Years of age 0 to 10	10 to 20	20 to 40	40 to 60	60+	Totals
Male	304	68	73	162	366	973
Female	187	37	47	116	221	608
Totals	491	105	120	278	587	1581

Assuming the model

$$\begin{matrix} p_1 & p_2 & p_3 & p_4 & p_5 & 1 \\ p_1 & p_2 & p_3 & p_4 & p_5 & 1, \end{matrix}$$

we have χ_4^2 equals 1.93, or nonsignificance. The opinion expressed as to what may be found in the data suggests that for an application of Neyman's $C(\alpha)$ technique we might choose for our alternate setup

$$\begin{matrix} p_1 + p_2\xi & p_2 + (p_3 - p_2)\xi & p_3(1 - \xi) & p_4(1 - \xi) & p_5 + p_4\xi & 1 \\ p_1 & p_2 & p_3 & p_4 & p_5 & 1 \end{matrix}$$

The test criterion is

$$T = \left(N_1 - \frac{NM_1}{M}\right)\frac{M_2}{M_1} + \left(N_2 - \frac{NM_2}{M}\right)\left(\frac{M_3 - M_2}{M_2}\right) - \left(N_3 - \frac{NM_3}{M}\right) - \left(N_4 - \frac{NM_4}{M}\right) + \frac{M_4}{M_5}\left(N_5 - \frac{NM_5}{M}\right). \tag{25}$$

Under H_0 the expected value of T is zero and

$$\operatorname{Var} T = \frac{nN}{M^2}\left[\frac{M_2^2}{M_1} + \frac{(M_3 - M_2)^2}{M_2} + M_3 + M_4 + \frac{M_4^2}{M_5}\right]. \tag{26}$$

From the table above, using $N = 973$, $n = 608$, we have

$$T = 13.06, \qquad \operatorname{Var} T = 131.18, \qquad \frac{T}{\sigma_T} = 1.14. \tag{27}$$

This is still not significant, so we would not reject the null hypothesis of no difference in the proportions.

REFERENCES

[1] J. NEYMAN, *Probability and Statistics; The Harald Cramér Volume*, New York, Wiley, 1959, pp. 213–234.

[2] ———, "Statistical problems in science. The symmetric test of a composite hypothesis," *J. Amer. Statist. Assoc.*, Vol. 64 (1969), pp. 1154–1171.

[3] R. A. FISHER, "The negative binomial distribution," *Ann. Eugenics*, Vol. 11 (1941), pp. 182–187.