# THE DERIVATION OF THE CHI-SQUARE TEST OF GOODNESS OF FIT

Jacqueline Prillaman

Submitted to McGill University in partial
fulfilment of the requirements for the
degree of Master of Science, August, 1956

I wish to express my gratitude to
Prof. W. Kozakiewicz for his generous
assistance and valuable guidance in
the preparation of this thesis.

TABLE OF CONTENTS

## Introduction

Let $x_1, x_2, \ldots, x_s$ be s independent sample values drawn from a population of normally distributed values with zero mean and unit variance. Then the variable

$$u = x_1^2 + x_2^2 + \ldots + x_s^2$$

is said to follow the $\chi^2$ distribution with parameter s. This parameter is called the number of degrees of freedom for reasons which will be explained later. Using the moment-generating function we shall prove that

$$(1) \qquad f(u) = \frac{1}{\Gamma\left(\frac{s}{2}\right)} \left(\frac{1}{2}\right)^{\frac{s}{2}} u^{\frac{s-2}{2}} e^{-\frac{u}{2}}, \qquad u > 0.$$

Since the basic variable, x, in the population has the frequency function

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

we have, for $\theta < \frac{1}{2}$,

$$M_{x_1^2}(\theta) = M_{x^2}(\theta) = \int_{-\infty}^{+\infty} e^{\theta x^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}(1-2\theta)} dx$$

$$= (1-2\theta)^{-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}} dy$$

$$= (1-2\theta)^{-\frac{1}{2}}.$$

And since the $x_i$ are independent, all have the same distribution as the variable x. We obtain

$$M_u(\theta) = M_{x_1^2 + \ldots + x_s^2}(\theta) = M_{x^2}^{s}(\theta) = (1-2\theta)^{-\frac{s}{2}}$$

by the well known property of moment-generating functions.

Now it can be seen that this is the moment-generating function of the distribution having the frequency function defined by (1). Indeed, the corresponding moment-generating function is given by the formula

$$M_u(\theta) = \int_0^\infty \frac{1}{\Gamma\left(\frac{s}{2}\right)} \left(\frac{1}{2}\right)^{\frac{s}{2}} u^{\frac{s-2}{2}} e^{-\frac{u}{2}} e^{\theta u}\, du$$

$$= \frac{1}{2^{\frac{s}{2}} \Gamma\left(\frac{s}{2}\right)} \int_0^\infty e^{-y} \left(\frac{2y}{1-2\theta}\right)^{\frac{s}{2}-1} \frac{2}{1-2\theta}\, dy$$

$$= \frac{(1-2\theta)^{-\frac{s}{2}}}{\Gamma\left(\frac{s}{2}\right)} \int_0^\infty e^{-y} y^{\frac{s}{2}-1}\, dy$$

$$= (1-2\theta)^{-\frac{s}{2}} .$$

Since a moment-generating function uniquely determines a distribution function, it follows that the distribution function of u is given by (1). Henceforth we shall denote $x_1^2 + x_2^2 + \ldots + x_s^2$ by $\chi^2$.

Let us now suppose that a population can be divided into k mutually exclusive classes and that $p_i$ is the proportion of individuals belonging to the ith class. Suppose that we select a sample of m individuals from that population and that the sample contains $n_1$ individuals from the first class, $n_2$ individuals from the second class, etc. Each variable $n_i$ follows the binomial distribution with mean $mp_i$ and standard deviation $\sqrt{mp_i q_i}$, where $q_i = 1 - p_i$. From the fact that the binomial distribution can be approximated well for large m by the normal distribution with mean $mp_i$ and standard deviation $\sqrt{mp_i q_i}$, it follows that the variable

$$\frac{n_i - mp_i}{\sqrt{mp_i q_i}}$$

has a distribution which approaches the normal distribution with mean zero and unit standard deviation as the size of the sample, m, becomes increasingly large. That is, each variable $\frac{n_i - mp_i}{\sqrt{mp_i q_i}}$ follows approximately the normal standard distribution. If we could assume for the moment that the $n_i$ are independent, then the variable

$$\frac{(n_1-mp_1)^2}{mp_1q_1} + \frac{(n_2-mp_2)^2}{mp_2q_2} + \cdot \cdot \cdot + \frac{(n_k-mp_k)^2}{mp_kq_k}$$

would follow approximately the $\chi^2$ distribution with k degrees of freedom. However, the $n_i$ are not completely independent, since in repeated sampling, their sum must always be m. Thus, if k – 1 of the $n_i$ are known, then the kth is necessarily already determined. As it turns out, by modifying the above expression by omitting the $q_i$ in the denominator, we obtain a variable which approximately follows the $\chi^2$ distribution with k – 1 degrees of freedom. It is obvious that if the $p_i$ are small, omitting the $q_i$ will not seriously alter the above expression.

Indeed the purpose of this paper is first to show that the expression

$$\sum_{1}^{k} \frac{(n_i - mp_i)^2}{mp_i} \tag{2}$$

follows the $\chi^2$ distribution with k – 1 degrees of freedom and furthermore to consider the distribution of (2) when the $p_i$ are not known and have to be estimated from the sample. To do this, we shall present two independent derivations of the Chi-Square test of goodness of fit. The first one, offered in Part One, is very intuitive but not completely rigorous. It uses geometrical arguments and simple approximation formulae. The second derivation, offered in Part Two, is based on the theory of definite positive quadratic forms and the theory of characteristic and moment-generating functions. However, the second derivation concerns only the case when the $p_i$ are known.

## Part One
## An Intuitive Approach to the Chi-Square Test of Goodness of Fit

### §I: Presentation of the Test

Let us suppose that we have a sample of m individuals which have been classified into k mutually exclusive classes, and that the observed frequency of the ith class is $n_i$. We wish to determine whether this sample could have been obtained by random sampling from a given parent population in which $p_i$ is the proportion of individuals belonging to the ith class. Since the sum of the theoretical frequencies must equal m, then $mp_i$ will be the theoretical frequency of the ith class.

Let us consider the variable

$$\chi^2 = \sum_{1}^{k} \frac{(n_i - mp_i)^2}{mp_i} \tag{3}$$

Obviously, $\chi^2$ is a measure of the compatibility between the sets of observed and theoretical frequencies. If the value of $\chi^2$ obtained is small, this would indicate near agreement between the observed and the expected frequencies while increasingly large values of $\chi^2$ would indicate increasingly poor agreement. If we could devise some test by which we can judge whether or not a specific $\chi^2$ indicates reasonable compatibility between the two sets of frequencies, we would then be in a position to determine whether or not our set of observed frequencies has been obtained in random sampling from the given population.

As an example, suppose that a die is tossed 24 times and that we set up a frequency distribution of the results. If the die is unbiased, each face has the probability 1/6 of occurring in a single roll. We would then get the following set of observed, ($n_i$), and theoretical, ($mp_i$), frequencies:

| Face: | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Observed: | 2 | 5 | 6 | 4 | 4 | 3 |
| Theoretical: | 4 | 4 | 4 | 4 | 4 | 4 |

Let us calculate $\chi^2$ as defined by (3).

$$\chi^2 = \frac{(2-4)^2}{4} + \frac{(5-4)^2}{4} + \frac{(6-4)^2}{4} + \frac{(4-4)^2}{4} + \frac{(4-4)^2}{4} + \frac{(3-4)^2}{4}$$

$$= 2.5$$

If we could determine that this value of $\chi^2$ showed reasonable compatibility between the two sets of frequencies, we could then assume that our set of observed data is not unusual and could have been obtained by rolling an unbiased die.

Let us imagine that we perform this experiment many times and calculate the $\chi^2$ value corresponding to each set of observed data. The $\chi^2$ values can take any value in the range $0 \leq \chi^2 < \infty$ Some of our values of $\chi^2$ will be smaller than 2.5, some larger. If we were to classify these values of $\chi^2$ into a relative frequency table, this table would tell us approximately into what percentage of such experiments various ranges of values of $\chi^2$ could be expected to be obtained. In particular we could determine what percentage of such experiments would give $\chi^2$ values greater than 2.5. If this percentage were large, we could assume that the sets of frequencies were reasonably compatible, i.e., our observed set was not unusual. If, however, that percentage were small, i.e., that there are hardly any other values of $\chi^2$ which are larger than our observed 2.5, we would conclude that the observed frequencies were not compatible with the frequencies expected for an unbiased die, and hence conclude that our die was biased. Thus we have for our "test of compatibility" the ratio of the number of all samples whose $\chi^2$ is greated than 2.5 to the total number of samples. We denote this ratio by P.

In the above discussion we have defined a certain ratio P by means of which we can test the compatibility between sets of observed and theoretical frequencies, i.e., a means by which we can determine the unusualness of our observed set as compared to the expected set, and we have proceeded empirically to determine the value of this ratio, P, by determining approximately the frequency distribution of $\chi^2$ for one particular problem. It is possible, however, to obtain an approximation to the frequency function of $\chi^2$ in the general case by theoretical methods. Indeed, we shall show that the frequency function defined by (1) is a close approximation to the frequency distribution of the $\chi^2$ given by formula (3) when m is large. Hence, we are able to determine values of P for all values of $\chi^2$ and these values of P are exactly what we find when we employ the $\chi^2$ tables.

## §II: The Number of Degrees of Freedom

Let us continue the example with the die, and determine the value of P from the $\chi^2$ table. To find the value of P corresponding to $\chi^2 = 2.5$ we need also to know the "number of degrees of freedom" which we will denote by s. The number of degrees of freedom, s, is defined to be

$$s = k - q$$

where k is the number of classes in the frequency distribution and q is the number of restrictions placed on the difference between the observed and theoretical frequencies, $(n_i - mp_i)$. In the problem of the die, there are six classes corresponding to the six faces on the die, hence $k = 6$. In determining the theoretical frequencies corresponding to each class it had to be assumed that the sum of the theoretical frequencies was equal to 24, the total of the observed frequencies. In effect, what we have assumed is that

$$\sum_{1}^{6} (n_i) = \sum_{1}^{6} (mp_i),$$

which transposed is

$$\sum_{1}^{6} (n_i - mp_i) = 0.$$

Thus, in this example, we have placed a restriction on the $(n_i - mp_i)$ that their sum from 1 to 6 must equal zero. Therefore the number of degrees of freedom in our problem is $s = 6 - 1 = 5$. Looking in the tables with $\chi^2 = 2.5$ and $s = 5$, we find $P = 0.77$, which tells us that 77% of all the other samples would have a $\chi^2 > 2.5$. Thus we can conclude that our sets of observed and theoretical frequencies are compatible and therefore our die was not biased.

In the example of the die, we knew 1/6 to be the probability for a given face to appear on a die from purely <u>a priori</u> considerations. We then obtained the theoretical frequencies for each class by determining the $mp_i$.

Suppose, however, that there were no <u>a priori</u> considerations given by which to determine the theoretical probabilities associated with each class and that these probabilities must be determined from the sample. This problem of determining the $p_i$ is usually done by the process of "fitting" a hypothetical distribution in the population to the observed data. Suppose, for example, that our sample has been derived from a supposedly normal population. Then to "fit" a normal curve to the data we find the sample mean and standard deviation, and use them as the corresponding parameters in the equation of the normal curve. We then can determine the $p_i$ by finding those areas under the normal curve which correspond to the class intervals of the sample data. We then test the goodness of fit of this curve to the observed data by calculating $\chi^2$ as defined in (3), establish what the number of degrees of freedom is, and read the value of P from the tables.

To determine the value of $\chi^2$ we must again assume that the sum of the theoretical frequencies is equal the total number of observed frequencies, i.e.

(4) $\sum_1^k (n_i - mp_i) = 0.$

It is evident that this is an assumption that is essential to the calculation of $\chi^2$ as defined by (3) in all cases. Hence, we shall always have lost at least one degree of freedom by the above restriction.

Let us examine what other restrictions we have placed on the $(n_i - mp_i)$ by making the sample mean and standard deviation serve as the population parameters. In determining the population mean we notice that approximately

$$\mu_1' = \sum_1^k a_i p_i = \frac{1}{m} \sum_1^k a_i (mp_i)$$

where $\mu_1'$ is the theoretical 1st moment about the origin, and $a_i$ is the class mark of the ith class. Note that $\mu_1'$ would be exactly equal to $\sum_1^k a_i p_i$ when the $p_i$ are given and the distribution is discrete. By demanding that the sample mean $\bar{x}$ serve as the population mean, we have

$$\frac{1}{m} \sum_1^k a_i n_i = \frac{1}{m} \sum_1^k a_i mp_i$$

i.e.,

(5) $$\sum_{1}^{k} a_i(n_i - mp_i) = 0.$$

Hence we have placed another restriction on the difference $(n_i - mp_i)$. Further, if we wish the standard deviation of the theoretical distribution to agree with that of the observed distribution, we should have, besides the condition $\bar{x} = \mu'_1$, which is equivalent to (5),

$$\mu'_2 = m'_2$$

where $\mu'_2$ is the theoretical second moment about the origin and

$$m'_2 = \frac{1}{m} \sum_{1}^{k} a_i^2 n_i .$$

Notice, however, that approximately

$$\mu'_2 = \sum_{1}^{k} a_i^2 p_i = \frac{1}{m} \sum_{1}^{k} a_i^2 mp_i .$$

This gives

(6) $$\sum_{1}^{k} a_i^2 (n_i - mp_i) = 0$$

In all, we have placed three restrictions on the $(n_i - mp_i)$. Hence, s in a problem of this type would then be k - 3.

Note that all three equations, (4), (5), and (6), are linear and homogeneous in $(n_i - mp_i)$. By extension, it can be seen that this is generally true of all restrictions where parameters of the theoretical distribution are derived from the observed data. Thus we see the precise relationship between degrees of freedom and linear, homogeneous restrictions in $(n_i - mp_i)$. The result is given in the following rule: The number of degrees of freedom, s, equals k - q, where k is the number of classes and q is the number of linear, homogeneous restrictions in $(n_i - mp_i)$ which arise in making parameters of the theoretical distribution agree with the parameters of the observed data.

Now we shall explain why the parameter s present in (1) is called the number of degrees of freedom. We shall prove in the next section that if the number of degrees of freedom is s, the variable defined by (3) follows the $\chi^2$ distribution with parameter s.

§III: Greenhood's Derivation of the $\chi^2$ Test of Goodness of Fit

We have said that a close approximation to the ratio P may be reached by theoretical methods. We now proceed with a theoretical approach to the determination of P. Let us suppose that we are given a set m of observed data $[v_1, v_2, \ldots, v_k]$ and that we wish to discover how unusual a sample our observed data is with respect to the theoretical frequencies $mp_1, mp_2, \ldots mp_k$, where the $p_i$ are known from <u>a priori</u> considerations. That is, we want to find P for the set $[v_1, \ldots, v_k]$. We know that the exact probability of getting the sample $[n_1, \ldots, n_k]$ is given by the multinomial expansion

$$(7) \qquad p[n_1 n_2 \ldots n_k] = p[n_k] = \frac{n!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k} .$$

If we now think of all the different possible samples $[n_1, \ldots, n_k]$ or $[n_k]$ for short, that might be drawn, we can find for each sample, from (7), the exact probability $p[n_k]$ associated with it. Obviously, unusual samples will have comparatively small probabilities, while usual samples will have larger values of $p[n_k]$.

We shall see that the value of $p[n_k]$ is connected with the value of $\chi^2$ as given by formula (3). [Compare formula (13).] It is intuitively obvious that the smaller the value of $p[n_k]$, the larger the value of $\chi^2$ . To determine the proportion of samples for which $\chi^2$ is larger than a given value $\chi_0^2$ we consider the specific sample $[v_k]$, with the corresponding value of $\chi^2$ equal to $\chi_0^2$ , and try to determine the proportion of samples whose probability of occurrence, $p[n_k]$, is less than $p[v_k]$ to the total number of samples.

Let us give concrete expression to this connection between values of $\chi^2$ and values of $p[n_k]$. Having determined P to be the ratio of samples whose probability of occurrence is less than $p[v_k]$ to the total number of samples, we need to examine the multinomial distribution function given in (7). We know that the point binomial

(8) $$\left\{p_1 + (1 - p_1)\right\}^m$$

gives the distribution of the probabilities associated with the number of observations falling into the ith class, and that function (7) is the combination of k distributions like (8), with each (8) along a different axis. Since it is impossible to picture (7) in the general case we shall content ourselves with picturing the three-dimensional frequency surface corresponding to $k = 2$. The three-dimensional surface corresponding to $k = 2$ resembles a mountain rising out of a plane. A plane parallel to the base plane intersects the frequency surface in a contour ellipse approximately. Were we to project the surface onto the plane, all the points that lie within the ellipse would have probabilities greater than points on the ellipse, while all the points lying outside the ellipse would have smaller probabilities. Letting the probability of the points lying on the ellipse be $p[v_1, v_2]$, then all points with smaller probabilities, $p[n_1, n_2] < p[v_1, v_2]$, lie outside the ellipse. Obviously, different-sized ellipses correspond to different given $p[v_1, v_2]$. We can, by an appropriate transformation, change the ellipses into circles with a common center lying directly beneath the peak of the frequency surface. By doing this, we can express the region where $p[n_1, n_2]$ is smaller than a certain value $p[v_1, v_2]$ as a function of the radius of the circle corresponding to $p[v_1, v_2]$.

By extension to the case of the arbitrary k, we have, instead of contour ellipses, ellipsoids. The transformation would give us hyperspheres. Different values of $p[v_k]$ now correspond to different layers of a k-dimensional ellipsoid. Let us denote the particular layer that corresponds to the $p[v_k]$ by $S_p$.

From this picture we now have the ratio P expressed by

$$(9) \qquad P = \frac{\sum p[n_i]}{\sum p[n_i]}$$

where the upper summation is over the outside of the surface $S_p$ and the lower summation is over the entire space. However, we should note here that both summations in (9) must be confined to points that satisfy the restriction imposed by (4) and other linear restrictions if they are present.

By Stirling's approximation for large factorials we have

$$p[n_k] = \frac{m!}{n_1!n_2!\ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

$$\cong \frac{m^{m+1/2} e^{-m} \sqrt{2\pi}\, p_1^{n_1+1/2} p_2^{n_2+1/2} \ldots p_k^{n_k+1/2}}{n_1^{n_1+1/2} n_2^{n_2+1/2} \ldots n_s^{n_s+1/2} e^{-n_1-n_2-\ldots-n_k} (2\pi)^{k/2} \sqrt{p_1 p_2 \ldots p_k}}$$

$$(10) \qquad \cong \frac{1}{(\sqrt{2\pi m})^{k-1} \sqrt{p_1 p_2 \ldots p_k}} \left(\frac{mp_1}{n_1}\right)^{n_1+1/2} \left(\frac{mp_2}{n_2}\right)^{n_2+1/2} \ldots \left(\frac{mp_k}{n_k}\right)^{n_k+1/2}$$

To make the transformation which will reduce our ellipsoids to a hypersphere with center at the origin, we use

$$n_i = mp_i + x_i\sqrt{mp_i}$$

or

$$(11) \quad x_i = \frac{n_i - mp_i}{\sqrt{mp_i}}$$

Substituting (11) for $n_i$ in the general term of (10), which is

$$\left(\frac{mp_i}{n_i}\right)^{n_i+1/2},$$

we get

$$\left(\frac{mp_i}{n_i}\right)^{n_i+1/2} = \left(\frac{mp_i}{mp_i+x_i\sqrt{mp_i}}\right)^{mp_i+x_i\sqrt{mp_i}+1/2}$$

$$= \left(\frac{mp_i+x_i\sqrt{mp_i}}{mp_i}\right)^{-mp_i-x_i\sqrt{mp_i}-1/2}$$

$$= \left(1 + \frac{x_i}{\sqrt{mp_i}}\right)^{-mp_i-x_i\sqrt{mp_i}-1/2}$$

as the general term which we shall call $h_i$.

By using the expansion

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$$

we perform the following transformation on $h_i$. We temporarily drop the subscript i for convenience of notation.

$$h = \left(1 + \frac{x}{\sqrt{mp}}\right)^{-mp-x\sqrt{mp}-1/2}$$

$$\log h = \log\left(1 + \frac{x}{\sqrt{mp}}\right)^{-mp-x\sqrt{mp}-1/2}.$$

$$-\log h = (mp + x\sqrt{mp} + 1/2)\log\left(1 + \frac{x}{\sqrt{mp}}\right)$$

$$= (mp + x\sqrt{mp} + 1/2)\left(\frac{x}{\sqrt{mp}} - \frac{x^2}{2mp} + \frac{x^3}{3mp^{3/2}} - \frac{x^4}{4(mp)^2} + \ldots\right)$$

$$= x\sqrt{mp} - \frac{x^2}{2} + \frac{x^3}{3\sqrt{mp}} - \frac{x^4}{4mp} + \ldots + x^2 - \frac{x^3}{2\sqrt{mp}}$$

$$+ \frac{x^4}{3mp} - \ldots + \frac{x}{2\sqrt{mp}} - \frac{x^2}{4mp} + \ldots .$$

$$- \log h = x\sqrt{mp} + \frac{x^2}{2} + B$$

where

$$B = \frac{x}{2\sqrt{mp}} - \frac{x^2}{4mp} + \ldots - \frac{x^3}{6\sqrt{mp}} + \frac{x^4}{12mp} - \ldots$$

We can neglect B since x is small with respect to $\sqrt{mp}$. Therefore

$$- \log h \cong x\sqrt{mp} + \frac{x^2}{2}$$

$$h \cong e^{-x\sqrt{mp}-1/2x^2} .$$

Substituting this result in (10) we get

$$(12) \qquad p[n_k] \cong \frac{1}{(\sqrt{2\pi m})^{k-1}\sqrt{p_1 p_2 \ldots p_k}} e^{-1/2(x_1^2+\ldots+x_k^2)} e^{-(x_1\sqrt{mp_1}+\ldots+x_k\sqrt{mp_k})}$$

From (11) we have

$$\sum_1^k x_i\sqrt{mp_i} = \sum_1^k (n_i - mp_i)$$

which equals zero by restriction (4). Thus (12) reduces to

$$p[n_k] \cong \frac{1}{(\sqrt{2\pi m})^{k-1}\sqrt{p_1 p_2 \ldots p_k}} e^{-\frac{r^2}{2}}$$

where

$$r^2 = x_1^2 + x_2^2 + \ldots + x_k^2 .$$

Furthermore, since

$$x_i^2 = \frac{(n_i - mp_i)^2}{mp_i}$$

it is evident that $r^2$ is the $\chi^2$ of the Chi-Square test defined in (3).

We have shown, through proper approximations and transformations, that $p[n_k]$ can be approximated well by a constant times $e^{-\frac{r^2}{2}}$, i.e.,

$$(13) \qquad p[n_k] \cong ce^{-\frac{r^2}{2}}$$

where $r^2$ is our familiar $\chi^2$ defined in (3). We also know that the volume over which we wish to sum, in the numerator of (9), is outside of the sphere $S_p$, whose radius is given by:

$$\chi_o = \sqrt{\frac{(v_1-mp_1)^2}{mp_1} + \ldots + \frac{(v_k-mp_k)^2}{mp_k}} \quad .$$

The last approximation in our proof is to replace the discrete summations in (9) by definite integrals. However, we have already noted that both summations in (9) must be confined to points that satisfy any restrictions that exist such as (4), (5) and (6). Let us investigate what effect such linear, homogeneous restrictions would have on the space over which we wish to integrate in order to determine P.

Let us begin by considering our three-dimensional frequency surface. The height of the ordinate erected at any point in the plane, say $(x_1,y_1)$, gives the probability that such a combination of events will occur. If we think of our point as being free to move over the x,y-plane, the height of the ordinate erected over the point increases or decreases as it moves, giving the exact probability for every point in the plane.

Now suppose that we are interested only in the probabilities of a certain set of the points on the plane, say those points for which the sum of the coordinates is 10. That is, we restrict our movement to only those points which lie on the line $x + y = 10$. Suppose that, in addition, we specify that we are interested only in those points which satisfy the condition $3x - y = 14$. Thus, we are only interested in the probability of the point $x = 6$, $y = 4$, or the intersection of the two lines $x + y = 10$ and $3x - y = 14$.

We note that we started with a plane over which we could move, but by imposing the linear restriction $x + y = 10$, our space was reduced to a line. Similarily, by imposing two linear restrictions, our space was restricted to a single point. We have, for every linear restriction, stepped down one dimension the space in which we can move. In general, this is true no matter how many dimensions has the space we are in. We have insisted that the restrictions be linear for the following reason. Linear equations are lines, planes, or hyperplanes while higher-powered equations are curves, curved surfaces and curved hypersurfaces. If we change from a three-space to a two-space curved surface, we have stepped down a dimension, but we are now moving over a highly complicated curved surface as compared with a flat plane. Recalling that the next step in our proof involves integration over a region similar to the one on which we are now moving, it is evident that if we are to integrate over curved hypersurfaces, we must use highly complicated line or surface integrals. Hence we insist on linear restrictions.

However, we have seen that restriction (4), and any others that might be present, are, in general, linear and homogeneous in $(n_i - mp_i)$. Let us see what effect the homogeneity has on the space over which we are moving. Again consider a sphere in three-dimensional space, radius r, equation $x^2 + y^2 + z^2 = r^2$. If we restrict our movement by some condition given as a linear homogeneous equation, we are in effect passing a plane through the origin which intersects our sphere in a circle $x'^2 + y'^2 = r^2$. In other words, we move down one dimension in the general family of hyperspheres, and still keep the same important constant r. If the restrictions had not been homogeneous, i.e., if the plane had not passed through the origin yet still intersected the sphere, the intersection would still have given a circle, but its radius would not have been the same r.

Generalizing the above, we have: The intersection of an n-dimensional hyperplane through the origin and an n-dimensional hypersphere with origin as center and radius r, is an n-1-dimensional hypersphere with radius r.

Thus we conclude that the effect of imposing q linear homoheneous restrictions in $(n_i - mp_i)$ on our summation in (9) is to restrict the space over which we can sum to $s = k - q$ dimensions, while at the same time, not distorting the important constant r.

Let us procede with the final step in our proof, i.e., replacing the discrete summations in (9) by definite integrals. From (13), the integral which we seek is of the form

$$(14) \qquad K\int_{r\geq\chi_0} e^{-\frac{r^2}{2}}\, dV$$

where dV is an element of volume in s-dimensional space. The constant K takes care of the constant multiplier in (13) plus any factor involved when we changed from an element of volume in the variables $[n_k]$ to an element of volume in $[x_k]$.

By replacing the sums in (9) by definite integrals over the proper limits (9) becomes

$$(15) \qquad P = \frac{\int_{r \geq \chi_0} e^{-\frac{r^2}{2}} dV}{\int_{r \geq 0} e^{-\frac{r^2}{2}} dV}$$

where the constant K has cancelled out and the integral in the denominator is over the entire s-dimensional space.

Let us see how our element of volume in the integration, dV, can be expressed more explicitly. As the integrand is a function of r alone, it would be appropriate to take dV as the volume between two hyperspheres of radius r and r + dr. In three space an element of volume, dV, between two spheres of radius r and r + dr would be $4\pi r^2 dr$, the surface of the sphere multiplied by dr. In two-dimensional space dV is given by $2\pi r dr$. In general, dV is a constant multiplied by r raised to a power one less than the number of dimensions of the hypersphere. Hence

$$(16) \qquad dV = Gr^{s-1}dr$$

Substituting (16) in (15) we get

$$(17) \qquad P = \frac{\int_{\chi_0}^{\infty} r^{s-1} e^{-\frac{r^2}{2}} dr}{\int_{0}^{\infty} r^{s-1} e^{-\frac{r^2}{2}} dr}$$

where, once again, the constant terms cancel out.

Let us calculate the value of the denominator. Using

$$\Gamma(m) = \int_0^{\infty} z^{m-1} e^{-z} dz$$

and letting $z = \frac{y^2}{2}$ , we have

$$\Gamma(m) = \int_0^\infty \left(\frac{y^2}{2}\right)^{m-1} e^{-\frac{y^2}{2}} y\,dy \quad .$$

Letting $m = \frac{s}{2}$ , then

$$\Gamma\left(\frac{s}{2}\right) = \left(\frac{1}{2}\right)^{\frac{s}{2}-1} \int_0^\infty (y^2)^{\frac{s}{2}-1} e^{-\frac{y^2}{2}} y\,dy$$

$$= \left(\frac{1}{2}\right)^{\frac{s}{2}-1} \int_0^\infty y^{s-2} e^{-\frac{y^2}{2}} y\,dy$$

$$= \frac{1}{2^{\frac{s-2}{2}}} \int_0^\infty y^{s-1} e^{-\frac{y^2}{2}} dy$$

Therefore

$$(18) \qquad \int_0^\infty r^{s-1} e^{-\frac{r^2}{2}} dr = (2)^{\frac{s-2}{2}} \Gamma\left(\frac{s}{2}\right)$$

Substituting (18) in (17) we get

$$P = \frac{1}{\Gamma\left(\frac{s}{2}\right)} \left(\frac{1}{2}\right)^{\frac{s-2}{2}} \int_{\chi_0}^\infty r^{s-1} e^{-\frac{r^2}{2}} dr \quad .$$

To arrive at our final frequency curve, the $\chi^2$ -curve, we substitute first $x = r^2$, which gives

$$P = \frac{1}{\Gamma\left(\frac{s}{2}\right)} \left(\frac{1}{2}\right)^{\frac{s}{2}} \left(\frac{1}{2}\right)^{-1} \int_{\chi_0^2}^\infty x^{\frac{s-1}{2}} e^{-\frac{x}{2}} \frac{1}{2} x^{-\frac{1}{2}} dx$$

$$(19) \qquad = \frac{1}{\Gamma\left(\frac{s}{2}\right)} \left(\frac{1}{2}\right)^{\frac{s}{2}} \int_{\chi_0^2}^\infty x^{\frac{s-2}{2}} e^{-\frac{x}{2}} dx \quad .$$

Therefore, we see that the proportion of samples having $\chi^2$ greater than a given value $\chi_0^2$ is obtained by the integration of the frequency function given in (1) from $\chi_0^2$ to $\infty$. Hence it is proved that the distribution of $\chi^2$, as defined in (3), is given by (1).

## Part Two

## A Rigorous Development of the Chi-Square Test of Goodness of Fit

### §I: The Normal Distribution in the Space of n-Dimensions

If

$$f(x_1,x_2,\ldots,x_n) = d\ e^{-\frac{1}{2}Q(x_1,x_2,\ldots,x_n)} ,$$

where $Q(x_1,\ldots,x_n)$ is a definite positive quadratic form of matrix A, then, with the proper choice of d, $f(x_1,\ldots,x_n)$ defines a frequency function in n-dimensions. The distribution with that frequency function is called the normal non-singular distribution.

We may write

$$(20)\qquad Q(x_1,\ldots,x_n) = X'AX = \sum_{i,j=1}^{n} A_{ij}x_ix_j \geq 0 ,$$

where

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and X' denotes, as usual, the transpose of the matrix X. It is well known that there exists an orthogonal transformation, X = CY, which reduces X'AX to the diagonal form

$$(21)\qquad Y'KY = \sum_{1}^{n} \kappa_i y_i^2 ,$$

where

$$K = \begin{pmatrix} \kappa_1 & & & 0 \\ & \kappa_2 & & \\ & & \ddots & \\ 0 & & & \kappa_n \end{pmatrix} .$$

If (20) is definite positive, then (21) will be definite positive, since that property is obviously invariant under any non-singular, linear transformation. It follows that all the $\kappa_i$ are positive.

Also, since

$$K = C'AC ,$$

then

$$|K| = |A| = \kappa_1 \kappa_2 \ldots \kappa_n > 0,$$

where, in a general way, we will denote the determinant of a matrix D by $|D|$.

To determine d, and for future purposes, it is convenient to find the moment-generating function corresponding to $f(x_1,\ldots,x_n)$, defined by

$$(22) \qquad \phi(t_1,\ldots,t_n) = d\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{T'X - \frac{1}{2}X'AX}\, dx_1\ldots dx_n \ .$$

To determine (22), we make the above considered substitution, $X = CY$, and also we replace the vector T with a new vector U by means of the contragredient substitution, $T = (C')^{-1}U$, which reduces to $T = CU$, since C is orthogonal. We obtain

$$T'X = (CU)'CY = U'(C'C)Y = U'Y$$

and, as we have seen,

$$X'AX = Y'KY \ .$$

Consequently

$$T'X - \frac{1}{2}X'AX = U'Y - \frac{1}{2}Y'KY$$

$$= u_1y_1 + \ldots + u_ny_n - \frac{1}{2}(\kappa_1y_1^2 + \ldots + \kappa_ny_n^2).$$

Hence the moment-generating function, with the t's expressed in terms of u's, becomes

$$\phi(t_1,\ldots,t_n) = d\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{\sum_1^n (u_jy_j - \frac{1}{2}\kappa_jy_j^2)}\, dy_1\ldots dy_n$$

$$= d\prod_1^n \int_{-\infty}^{\infty} e^{u_jy_j - \frac{1}{2}\kappa_jy_j^2}\, dy_j$$

$$= d \prod_{1}^{n} e^{\frac{u_j^2}{2K_j}} \sqrt{\frac{2\pi}{K_j}}$$

$$= d \frac{(2\pi)^{\frac{n}{2}} e^{\sum_{1}^{n} \frac{u_j}{2K_j}}}{\sqrt{K_1 \cdots K_n}}$$

since, for $a > 0$,

$$\int_{-\infty}^{\infty} e^{bx - \frac{1}{2} ax^2} dx = e^{\frac{b^2}{2a}} \sqrt{\frac{2\pi}{a}} .$$

Now

$$\sum_{1}^{n} \frac{u_j^2}{K_j} = \sum_{1}^{n} K_j^{-1} u_j^2$$

$$= U'K^{-1}U = U'(C'AC)^{-1}U$$

$$= U'(C^{-1}A^{-1}C)U$$

$$= U'(C'A^{-1}C)U$$

$$= (CU)'A^{-1}CU$$

$$= T'A^{-1}T .$$

Hence we get

$$\phi(t_1, \ldots, t_n) = d \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{|A|}} e^{\frac{1}{2} T'A^{-1}T} .$$

Substituting $T = 0$, we obtain

$$\phi(0, \ldots, 0) = \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{|A|}} d = 1 ,$$

hence

$$d = \frac{\sqrt{|A|}}{(2\pi)^{\frac{n}{2}}} .$$

Therefore

$$(23) \qquad f(x_1,\ldots,x_n) = \frac{\sqrt{|A|}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} Q(x_1,\ldots,x_n)}$$

$$= \frac{\sqrt{|A|}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} X'AX}$$

and

$$(24) \qquad \phi(t_1,\ldots,t_n) = e^{\frac{1}{2} T'A^{-1}T} \quad .$$

We shall prove now that if the variables $x_1,\ldots,x_n$ follow a non-singular normal distribution with the frequency function

$$f(x_1,\ldots,x_n) = d\ e^{-\frac{1}{2}\sum_{i,j=1}^{n} A_{ij}x_ix_j} \ ,$$

$$d = \frac{\sqrt{|A|}}{(2\pi)^{\frac{n}{2}}} \ ,$$

then the expression

$$Q(x_1,\ldots,x_n) = \sum_{i,j=1}^{n} A_{ij}x_ix_j$$

follows the Chi-Square distribution with n degrees of freedom.

The moment-generating function of $Q(x_1,\ldots,x_n)$ is given by

$$(25) \qquad \phi(\theta) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{Q\theta}\ d\ e^{-\frac{1}{2}Q}\ dx_1\ldots dx_n$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} de^{-\frac{1}{2}(1-2\theta)Q}\ dx_1\ldots dx_n$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} d\ e^{-\frac{1}{2}\sum_{i,j=1}^{n} A_{ij}(1-2\theta)x_ix_j}\ dx_1\ldots dx_n \ .$$

For $\theta < \frac{1}{2}$, $\sum_{i,j}^{n} A_{ij}\ (1-2\theta)x_ix_j$ is a definite positive quadratic form.

It has been found, in the determination of the constant d, that

$$\frac{\sqrt{|A|}}{(2\pi)^{\frac{n}{2}}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}\sum_{i,j=1}^{n} A_{ij}x_i x_j} \, dx_1 \ldots dx_n = 1 \quad .$$

Hence

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}\sum_{i,j=1}^{n} A_{ij}x_i x_j} \, dx_1 \ldots dx_n = \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{|A|}} \quad .$$

Consequently

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}\sum_{i,j=1}^{n} A_{ij}x_i x_j (1-2\theta)} \, dx_1 \ldots dx_n = \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{|\bar{A}|}}$$

where

$$\bar{A} = (1-2\theta)A$$

and

$$|\bar{A}| = (1-2\theta)^n |A| \quad .$$

Substituting this result in (25), we find

$$\phi(\theta) = (1 - 2\theta)^{-\frac{n}{2}} \quad ,$$

which is the moment-generating function of the Chi-Square distribution with n degrees of freedom. Hence the theorem is proved.

## §II: The Limiting Distribution of the Chi-Square Variable by the Method of Characteristic Functions

Suppose we have, as before, a population in which each element belongs to one and only one of the classes $C_1, C_2, \ldots, C_k$. Let $p_1, p_2, \ldots, p_k$, $\sum_{i=1}^{k} p_i = 1$, be the probabilities associated with $C_1, C_2, \ldots, C_k$ respectively. In a sample of size n, let $n_1, \ldots, n_k$

be the numbers of elements falling into $C_1, \ldots, C_k$ respectively. We have seen that the probability law of the $n_i$ is given by

$$\frac{n!}{n_1! \ldots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k} \quad .$$

We have $E(n_i) = np_i$. In view of the Central Limit Theorem, it is clear that the limiting distribution, as $n \to \infty$, of each of the quantities

$$\frac{\left(\frac{n_i}{n} - p_i\right)\sqrt{n}}{\sqrt{p_i(1-p_i)}} \quad , \qquad i = 1, \ldots, k$$

is the normal distribution with mean 0 and standard deviation 1. We shall now investigate the limiting joint distribution of the set

$$x_i = \frac{(n_i - np_i)}{\sqrt{n}} \quad , \quad i = 1, \ldots, k \quad .$$

Since $\sum_1^k x_i = 0$, only $k - 1$ of the $x_i$ are functionally independent. It is sufficient to consider the limiting joint distribution of the first $k - 1$ of the $x_i$.

We know that the moment-generating function of $x_1, \ldots, x_{k-1}$, say $\phi(\theta_1, \ldots, \theta_{k-1})$, is equal to $\Phi(\theta_1, \ldots, \theta_{k-1}, 0)$, where $\Phi(\theta_1, \ldots, \theta_k)$ is the moment-generating function of $x_1, \ldots, x_k$. However

$$\Phi(\theta_1, \ldots, \theta_k) = E(e^{\sum_1^k \theta_i x_i})$$

$$= e^{-\sqrt{n}\sum_1^k \theta_i p_i} \sum_1^k \frac{n!}{n_1! \ldots n_k!} \prod_1^k \left(p_i e^{\frac{\theta_i}{\sqrt{n}}}\right)^{n_i}$$

$$= e^{-\sqrt{n}\sum_1^k \theta_i p_i} \left(p_1 e^{\frac{\theta_1}{\sqrt{n}}} + p_2 e^{\frac{\theta_2}{\sqrt{n}}} + \ldots + p_k e^{\frac{\theta_k}{\sqrt{n}}}\right)^n .$$

Consequently

$$(26) \qquad \phi(\theta_1,\ldots,\theta_{k-1}) = e^{-\sqrt{n}\sum_{1}^{k-1}\theta_i p_i}\left(p_1 e^{\frac{\theta_1}{\sqrt{n}}} + \ldots + p_{k-1}e^{\frac{\theta_{k-1}}{\sqrt{n}}} + p_k\right)^n .$$

Expanding each of the exponentials in (26) and taking logarithms, we have

$$(27) \qquad \log \phi = -\sqrt{n}\sum_{1}^{k-1}\theta_i p_i + n\log\left[1 + \sum_{1}^{k-1}\frac{\theta_i p_i}{\sqrt{n}} + \sum_{1}^{k-1}\frac{\theta_i^2 p_i}{2n} + O\left(\frac{1}{n^{3/2}}\right)\right] .$$

Noticing that

$$\log(1+x) = x - \frac{x^2}{2} + |R(x)|$$

where

$$|R(x)| = \frac{|x|^3}{3} + \frac{|x|^4}{4} + \ldots$$

$$\leq \frac{|x|^3}{3} + \frac{|x|^4}{3} + \ldots$$

$$= \frac{1}{3}\frac{|x|^3}{1-|x|} \leq \frac{2}{3}|x|^3 \qquad \text{for } |x| \leq \frac{1}{2} ,$$

we find that

$$\log(1+x) = x - \frac{x^2}{2} + \alpha x^3 , \qquad |\alpha| < 1 \text{ for } |x| \leq \frac{1}{2} .$$

Hence we get from (27)

$$\log \phi = \sum_{1}^{k-1}\frac{\theta_i^2 p_i}{2} - \sum_{i,j=1}^{k-1}\frac{\theta_i\theta_j p_i p_j}{2} + O\left(\frac{1}{\sqrt{n}}\right) .$$

Therefore we have

$$\lim_{n\to\infty} \phi = e^{\frac{1}{2}\sum_{i,j=1}^{k-1} A^{ij}\theta_i\theta_j}$$

where $A^{ij} = p_i\partial_{ij} - p_i p_j$, $i,j = 1,\ldots,k-1$, and $\partial_{ij} = 1,\ i=j$ 
$= 0,\ i\neq j$ .

We shall prove that $\sum_{i,j=1}^{k-1} A^{ij}\theta_i\theta_j$ is a definite positive quadratic form. Considering

$$\sum_{i,j=1}^{k-1} A^{ij}\theta_i\theta_j = \sum_{1}^{k-1} \theta_j^2 p_j - (\sum_{1}^{k-1} \theta_j p_j)^2$$

and letting

$$\tilde{\theta} = \sum_{1}^{k-1} \theta_j p_j$$

we have

$$(28) \quad \sum_{i,j=1}^{k-1} A^{ij}\theta_i\theta_j = \sum_{1}^{k-1} p_j(\theta_j - \tilde{\theta})^2 + p_k\tilde{\theta}^2$$

In this form it is evident that $\sum_{i,j=1}^{k-1} A^{ij}\theta_i\theta_j$ is non-negative and equal to zero if and only if $\tilde{\theta} = 0$ and $\theta_j = \tilde{\theta}$, i.e., if $\theta_j = 0$ for $j = 1,\ldots,k-1$.

It follows, therefore, from the previous discussion that the limiting frequency function for the joint distribution of the $x_i$ is given by

$$(29) \quad \frac{\sqrt{|A|}}{(2\pi)^{\frac{k-1}{2}}} \; e^{-\frac{1}{2}\sum_{i,j=1}^{k-1} A_{ij}x_i x_j}$$

where $A_{ij} = (A^{ij})^{-1}$ .

It may be verified readily, by the multiplication of matrices, that

$$A_{ij} = \frac{\partial_{ij}}{p_i} + \frac{1}{p_k}$$

and therefore

$$(30) \quad \sum_{i,j=1}^{k-1} A_{ij}x_i x_j = \sum_{1}^{k-1} \frac{x_i^2}{p_i} + \frac{1}{p_k} (\sum_{1}^{k-1} x_i)^2 \quad .$$

We have seen that if $x_1,\ldots,x_{k-1}$ are random variables having distribution (29) then $\sum_{i,j=1}^{k-1} A_{ij}x_i x_j$ is distributed according to the Chi-Square law with k-1 degrees of freedom.

We now replace $x_i$ by $(n_i - np_i)/\sqrt{n}$ in (30), denoting the result by $\chi^2$ , and we obtain the familiar expression of $\chi^2$ as seen before in (3), i.e.

$$(31) \quad \chi^2 = \sum_{i}^{k} \frac{(n_i - np_i)^2}{np_i} \quad .$$

We wish to conclude that the limiting distribution of $\chi^2$ is identical with the distribution of $\sum_{i,j=1}^{k} A_{ij}x_ix_j$ where the $x_i$ are distributed according to (29). That is, the limiting distribution of the expression in (31) is the Chi-Square distribution with k-1 degrees of freedom. To do this we prove the following theorem.

Consider the random vector variables $(X_1^{(1)}, X_2^{(1)}, \ldots, X_p^{(1)})$, $(X_1^{(2)}, X_2^{(2)}, \ldots, X_p^{(2)})$, . . . , $(X_1, X_2, \ldots, X_p)$, with the probability functions $P_1(S), P_2(S), \ldots, P(S)$ and the distribution functions $F_1(x_1,x_2,\ldots,x_p)$, $F_2(x_1,x_2,\ldots,x_p)$, . . . , $F(x_1,x_2,\ldots,x_p)$ respectively. (We use here Cramer's terminology.) Assume that the sequence $F_n(x_1,x_2,\ldots,x_p)$ converges to $F(x_1,x_2,\ldots,x_p)$ in all non-excluded points of the latter.[1) ] Let $f(x_1,\ldots,x_p)$ be a function which is continuous and defined everywhere in the p-dimensional space. Then the distribution function of the variable $f(X_1^{(n)}, X_2^{(n)}, \ldots, X_p^{(n)})$ converges to the distribution function of $f(X_1, X_2, \ldots, X_p)$.

We shall give the proof only in the two-dimensional case as it is technically simpler. However, generalization to any finite number of dimensions is immediate.

We shall complete the proof by using the method of characteristic

1) See Cramer, (Ref. 1), p. 83.

functions. It is sufficient to prove that the characteristic function of $f(X_1^{(n)}, X_2^{(n)})$, say $\psi_n(t)$, converges to that of $f(X_1, X_2)$, say $\psi(t)$.

We have

$$\psi_n(t) = \int_{R_2} e^{if(x_1,x_2)t}\, dP_n = \int_{R_2}\!\!\int e^{if(x_1,x_2)t}\, d^2F_n$$

$$\psi(t) = \int_{R_2} e^{if(x_1,x_2)t}\, dP = \int_{R_2}\!\!\int e^{if(x_1,x_2)t}\, d^2F \quad .$$

Let K be a continuity rectangle of $F(x_1,x_2)$. Then by the Helly-Bray convergence theorem[2] we have

$$(32) \qquad \lim_{n\to\infty} \int_K e^{if(x_1,x_2)t}\, dP_n = \int_K\!\!\int e^{if(x_1,x_2)t}\, dP$$

for every t.

Let $\epsilon$ be an arbitrary positive number. The continuity rectangle K can be chosen such that

$$P(K) > 1 - \epsilon \quad .$$

Representing P(K) and $P_n(K)$ as the 2nd difference it follows that

$$\lim_{n\to\infty} P_n(K) = P(K).$$

Consequently there exists a number $n_0$ such that

$$P_n(K) > 1 - 2\epsilon$$

for $n > n_0$.

However

$$\psi_n(t) = \int_K e^{if(x_1,x_2)t}\, dP_n + \int_{K*} e^{if(x_1,x_2)t}\, dP_n$$

$$\psi(t) = \int_K e^{if(x_1,x_2)t}\, dP + \int_{K*} e^{if(x_1,x_2)t}\, dP$$

where $K* = R_2 - K$, and

2) See Cramer, (Ref. 1), p. 74.

$$\left| \int_{K^*} e^{if(x_1,x_2)t} \, dP_n \right| \leq \int_{K^*} dP_n = P_n(K^*) = 1 - P_n(K) < 2\epsilon$$

for $n > n_0$. Similarily

$$\left| \int_{K^*} e^{if(x_1,x_2)t} \, dP \right| \leq \int_{K^*} dP = P(K^*) = 1 - P(K) < \epsilon \quad .$$

By (32), for all sufficiently large n

$$\left| \int_{K} e^{if(x_1,x_2)t} \, dP_n - \int_{K} e^{if(x_1,x_2)t} \, dP \right| < \epsilon \quad .$$

Consequently, for all sufficiently large n, we ha e

$$\left| \psi_n(t) - \psi(t) \right| < 4\epsilon$$

which proves that $\psi_n(t)$ converges to $\psi(t)$.

The problem of the limiting Chi-Square distribution in the case where the probabilities are not known and have to be estimated from the sample can be treated in a similar way. However, the proof then is much more difficult, requiring some special techniques, and will be omitted here. The interested reader is referred to texts by Cramer, (Ref. 1), pp. 426-434, and Wilkes, (Ref. 5), pp. 219-220.

# BIBLIOGRAPHY

1. Cramer, Harald: Mathematical Methods of Statistics. Princeton University Press, Princeton, 1946.

2. Cramer, Harald: The Elements of Probability Theory. John Wiley and Sons, New York, 1955.

3. Greenhood, E. Russell: A Detailed Proof of the Chi-Square Test of Goodness of Fit. Harvard University Press, Cambridge, 1940.

4. Hoel, Paul G.: Introduction to Mathematical Statistics. John Wiley and Sons, New York, 1947.

5. Wilkes, S. S.: Mathematical Statistics. Princeton University Press, Princeton, 1946.