Experiment 212: Central Limit Theorem

Aims

The primary aim of this experiment is to test physically the Central Limit Theorem of statistics. As well there are four secondary aims:

1. To give exercise in deriving an experimental distribution.
2. To emphasise the Gaussian distribution because it is the simplest and the most used.
3. To emphasise the Central Limit Theorem since this is the reason that the Gaussian turns up so often in physics.
4. To show how a sample approaches the theoretical curve as the sample number is increased.

Reference

1. Eton Statistical and Math Tables, Heinemann.

Introduction

The Central Limit Theorem states that:

If $x = x_1 + x_2 + x_3 + \ldots + x_r$ where $x_1, x_2, x_3, \ldots, x_r$ are independent variables each with its own finite probability distribution, then the probability distribution of x tends towards a theoretical probability distribution known as the Gaussian (or "Normal") distribution as the number of independent variables r included in x is increased (see Figure 1).

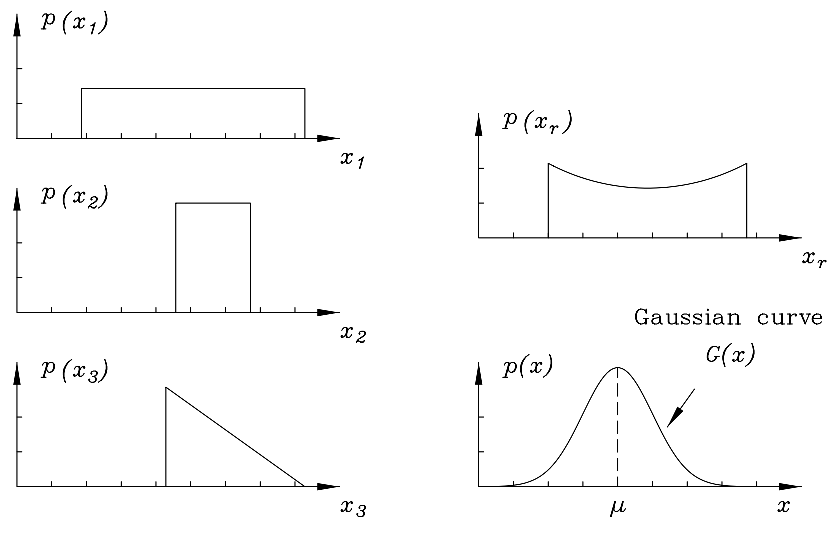

Figure 1:

The Gaussian distribution $G(x)$ has the equation:

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{1}$$

where μ is the population mean and σ is the population standard deviation.

Some measurements have errors introduced by several causes (parallax, vibration, friction, mental uncertainty, etc.) and so the resulting distribution will be similar to the Gaussian distribution. This is one reason for the frequent appearance of the Gaussian distribution in physics (see Appendix A). The tendency for $p(x)$ to become $G(x)$ is very rapid. Usually only 5 - 10 independent variables are required for $p(x)$ to be very close to $G(x)$.

We can try to simulate the effect of many contributory errors by using "loaded" dice. Each of the dice represents a single random variable and will have a distinctive (non-Gaussian) distribution. The supplied dice are designed to have irregular probability distributions (such as those shown in Figure 2) to show how general the Central Limit Theorem is. The tendency to the Gaussian is even more rapid for smooth distributions with central peaks than it is for irregular skew distributions.

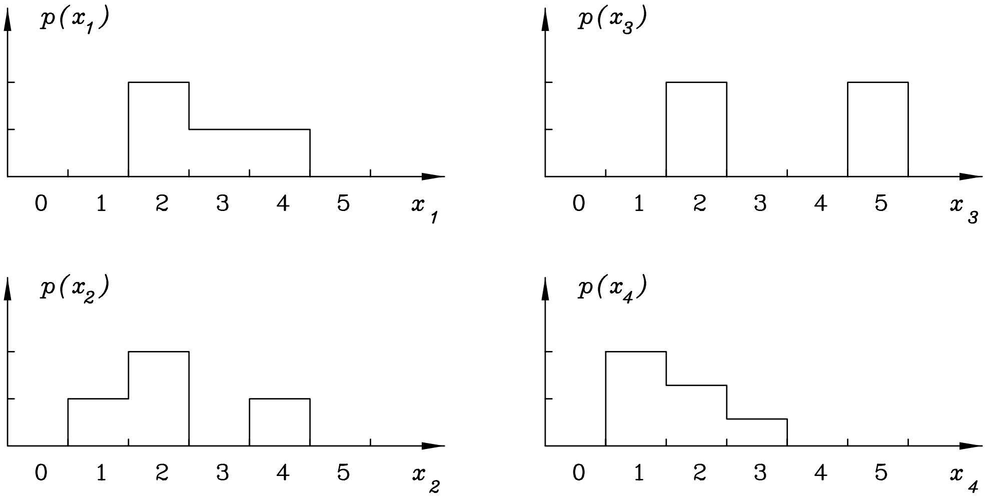

Figure 2:

Procedure

(1) Select 3 dice from those supplied.

(2) Examine the selected dice and notice whether their distributions will be irregular or not. (You may use the most irregular and lopsided or the most uniform with a central peak - please yourself).

(3) Draw (roughly) the distribution that you would expect from each selected die.

(4) Toss the selected dice 400 times, and use for your variable x, the sum of the numbers on the downturned faces. (This is fairly tiring but two students taking turns should complete this in less than one laboratory period).

(5) Select 6 more dice and add these to your previous 3 to give you a total of 9 dice.

(6) Repeat (2) to (4) with this larger number of dice.

(7) Calculate the (sample) mean m of each set of results and use this as your best estimate of μ, the population mean.

(8) Calculate σ, the standard deviation of each set, i.e. your best estimate of the standard deviation of the population, using the equation:

$$\sigma = \sqrt{\frac{\sum_x f(x)(x-m)^2}{n-1}} \tag{2}$$

where $f(x)$ = number of times the sum of x numbers is obtained, n = number times the selected dice is tossed and the summation extends over all possible sums which can be obtained using the chosen dice.

(9) Draw a histogram of the experimental results for 3 dice. On the same axes, draw a curve showing the frequencies expected assuming a Gaussian distribution with mean and standard deviation as calculated in (7) and (8). (See also Figure 3.)

(10) Repeat procedure (9) for the nine dice.

(11) Apply the χ^2 test to the two sets to test whether they agree with the corresponding Gaussians. (See Appendix B).

Question:

1. Do you consider the theorem to have been verified?

APPENDIX A

The Role of the Gaussian (or Normal) Distribution in Statistics

The normal distribution was first found in 1733 by De Moivre in connection with his discussion of the limiting form of the binomial distribution. De Moivre's discovery seems, however, to have passed unnoticed, and it was not until long afterwards that the normal distribution was rediscovered by Gauss and Laplace. The latter did, in fact, touch on the subject in some papers about 1780, though he did not go deeper into it before his great work of 1812. Gauss and Laplace were both led to the normal function in connection with their work on the theory of errors of observation. Laplace gave, moreover, the first (incomplete) statement of the general theorem studied above under the name of the Central Limit Theorem, and made a great number of important applications of the normal distribution to various questions in the theory of probability.

Under the influence of the great works of Gauss and Laplace, it was for a long time regarded as an axiom that statistical distributions of practically all kinds would approach the normal distribution as an ideal limiting form, if only we could dispose a sufficiently large number of sufficiently accurate observations. The deviation of any random variable from its mean was regarded as an "error", subject to the "law of errors" expressed by the normal distribution.

Even though this view was definitely exaggerated and has had to be considerably modified, it is undeniable that, in a large number of important applications, we meet distributions which are at least approximately normal. Such is the case, for example, with the distributions of errors of physical and astronomical measurements, a great number of demographical and biological distributions, and many others.

The central limit theorem affords a theoretical explanation of these empirical facts. According to the "hypothesis of elementary errors" introduced by Hagen and Bessel, the total error committed at a physical or astronomical measurement is regarded as the sum of a large number of mutually independent elementary errors. By the central limit theorem, the total error should then be approximately normally distributed. In a similar way, it often seems reasonable to regard an observed random variable, (e.g. in some biological investigation) as being the total effect of a large number of independent causes which sum up their effects. The same point of view may be applied to the variables occurring in many technical and economical questions. Thus the total consumption of electric energy delivered by a certain producer is the sum of the quantities consumed by the various customers, the total gain or loss on the risk business of an insurance company is the sum of the gains or losses on each single policy, etc.

In cases of this character, we should expect to find at least approximately normal distributions. If the number of components is not sufficiently large, or if the various components cannot be regarded as strictly additive and independent, modifications of the central limit theorem may still show that the distribution is approximately normal, or they may indicate the use of some distribution closely related to the normal, such as the asymptotic expansion or the logarithmico-normal distribution.

Under the conditions of the central limit theorem, the arithmetic mean of a large number of independent variables is approximately normally distributed. Furthermore, this property holds true even for certain functions of a more general character than the mean. These properties are of a fundamental importance for many methods used in statistical practice, where we are largely concerned with means and other similar functions of the observed values of random variables.

There is a famous remark by Lippman (quoted by Poincare), to the effect that "everybody believes in the law of errors, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact". It seems appropriate to comment that both parties are perfectly right, provided that their belief is not too absolute; mathematical proof tells us that, under certain qualifying conditions, we are justified in expecting a normal distribution, while statistical experience shows that, in fact, distributions are often approximately normal.

APPENDIX B

χ^2 Test

Comparison of Sample with Population

It is often important to test a theory or hypothesis which indicates a particular population distribution by comparing it with a series of practical results (a sample). For example a physicist wishing to test whether the counts given by a nuclear device were random, might compare the distribution of counts/sec obtained experimentally with the Poisson distribution.

Owing to random fluctuations, the sample histogram will not agree completely with the theoretical distribution. We would expect large relative deviations for small samples and small relative deviations for large samples. We require a convenient measure of the deviation of the sample histogram from the hypothetical distribution (see Figure 3). The χ^2 test provides this.

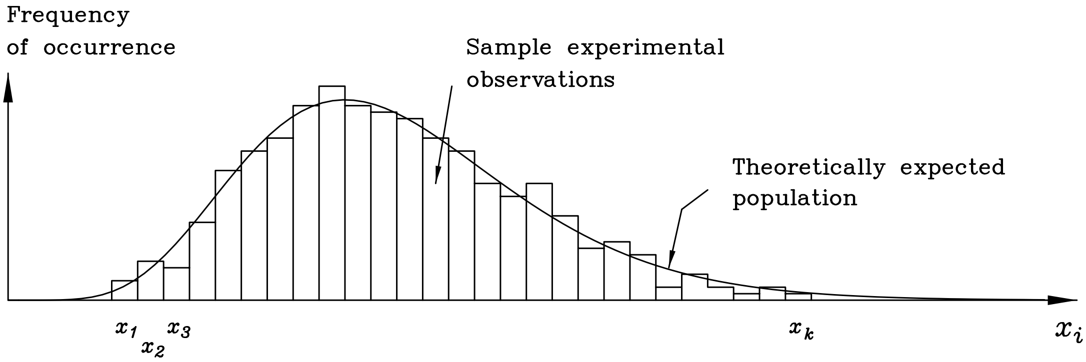

Figure 3:

Suppose $x_1, x_2, x_3, \ldots, x_i, \ldots, x_k$ are the k different results, with observed frequencies $O(x_i)$ (called $f(x)$ in this experiment) and expected frequencies $E(x)$, then we define χ^2 as:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left[O(x_i) - E(x_i)\right]^2}{E(x_i)} \tag{3}$$

The theory of χ^2 requires that each $E(x_i)$ be greater than about 5.

Histogram "bins" can be grouped to ensure that $E(x_i) > 5$, e.g. if:

x_i	...	11	12	13	14	15	16	17
$E(x_i)$	...	11.2	5.3	2.5	1.2	0.5	0.2	0.1
$O(x_i)$	...	9	6	1	0	1	0	0

then these results should be grouped to give

x_i	...	11	12 – 17
$E(x_i)$	...	11.2	9.8
$O(x_i)$	...	9	8

Interpretation

If χ^2 is very small: the experimental results fit the theoretical almost exactly. This is suspicious — there should be some random scatter.

If χ^2 is moderate: the experimental results do not deviate very much from the theoretical. This is reassuring.

If χ^2 is very large: the experimental results disagree with theoretical predictions. This is alarming.

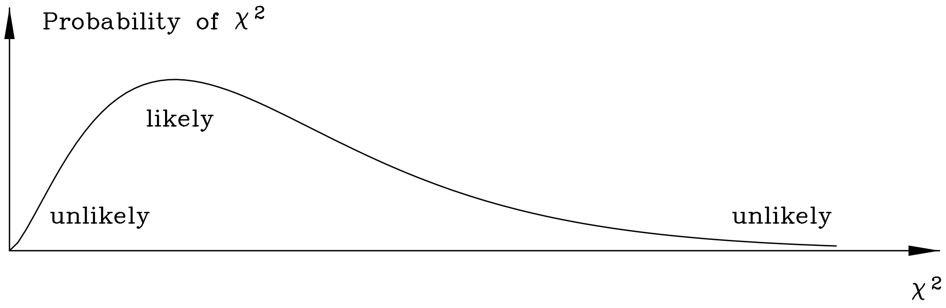

Figure 4:

The distribution of χ^2 (drawn in Figure 4) will depend upon two things:

(i) The number of results k, since χ^2 is likely to be larger if there are more points to be compared;

(ii) The number of parameters which have been chosen to help the theoretical curve fit the experimental. If the mean or standard deviation has been chosen to be the same as the experimental then we should expect a better fit (i.e. a lower χ^2).

These are combined in the concept of the number of degrees of freedom "ν", which refers to the number of independent points in the histogram (after grouping) — i.e. the number of independent terms in the sum (3) (after grouping). Thus:

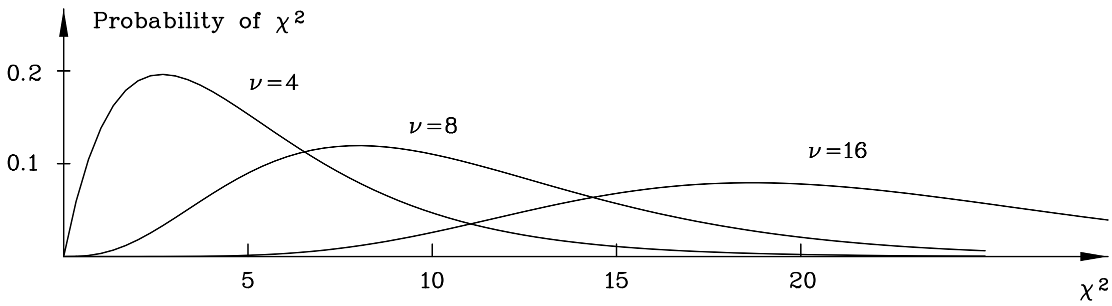

Figure 5:

(a) If the area under the theoretical curve is made equal to the area under the experimental curve, then

$$\nu = k - 1$$

(b) If the area and the mean of the theoretical curve are made to fit that of the experimental curve, then

$$\nu = k - 2$$

(c) If the area, the mean and the standard deviation of the theoretical curve are made to fit those of the experimental curve (as is the case in this experiment), then

$$\nu = k - 3$$

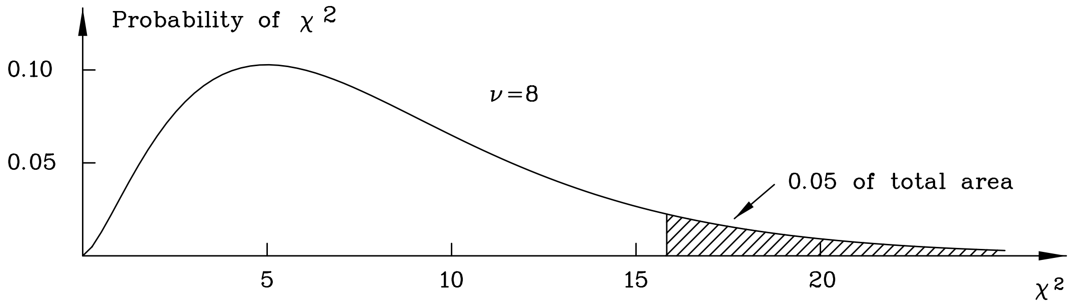

Figure 6:

Tables of the integral of χ^2 are readily available (Eton Statistical and Math Tables). In these tables, the probability P is that of obtaining values greater than χ^2. Thus if your value of χ^2 is greater than the χ^2 which has $P = 0.05$ then it must lie in the shaded area of Figure 6, and is unlikely.

Similarly, if your value of χ^2 is less than the χ^2 which has $P = 0.95$ then it must lie in the unshaded area of Figure 7 and is again unlikely.

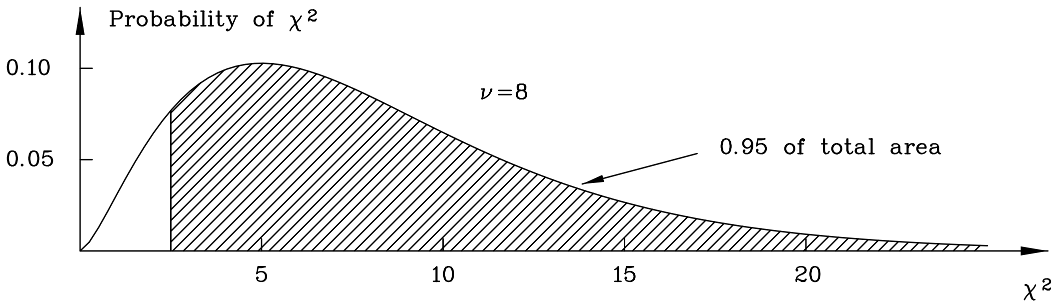

Figure 7:

However, if your value of χ^2 is between the values of χ^2 for $P = 0.05$ and $P = 0.95$ then we usually assume that the experiment has not disagreed with the theory.

Example

If your value of χ^2 is between 7.26 and 24.99 for 15 degrees of freedom then your experiment has not disagreed with the theory.

Write-up

The write-up should cover the following:

1. Draw (roughly) the distribution that you would expect from each selected die.
2. Calculate the mean of both sets of results.
3. Calculate the standard deviation of both sets.
4. Draw Gaussian curves with the same means and standard deviations as your two sets of measurements.
5. Draw histograms of the two experimental sets each on the same sheet as the corresponding Gaussian curve.

6. Calculate χ^2 for the two sets.
7. Decide whether or not the theorem has been verified.

List of Equipment

1. Plastic Cup.
2. 9 Assorted Dice

A. Chisholm

Z.C. Tan

Last update: May 15, 2008