APPENDIX G

# Statistical Tests of Hypotheses

In our work on the generation of pseudorandom numbers in Appendix F we considered two of the tests of the quality of these numbers, and we encountered possible deviations from theoretical expectations based on the assumption of complete randomness. Similarly, in our examples involving the transformation and rejection methods for nonuniform pseudorandom numbers, we noted the statistical fluctuations associated with finite sequences of random numbers. One encounters such fluctuations often in dealing with stochastic processes, including any physical measurements (and not just with random number generators), so it is important to consider them in a little more detail. For example, how can we tell if these fluctuations are too large or too small? That is, how do we know that they are just the expected statistical fluctuations?

This question[1] is a central issue in statistics and is discussed at great length in many textbooks (see, for example, Press et al. [1986]). The standard statistical test for determining if an observed distribution is consistent with a model (that is, theoretical) distribution is based on what is known as the *chi-square statistic*. This is defined by first separating the observed values into suitable bins, and then comparing the number of times the measured value falls into each bin to the statistical expectations. A quantity called $\chi^2$ is defined by

$$\chi^2 = \sum_i \frac{(N_i - n_{\text{ideal}})^2}{n_{\text{ideal}}} . \tag{G.1}$$

Here $N_i$ is the number of events that are *measured* to fall into bin $i$, and $n_{\text{ideal}}$ is the number of events that the theoretical model *predicts* should fall into each bin.[2] The quantity $\chi^2$ is thus a quantitative measure of how much the observed distribution differs from the theoretical one. We should not expect this difference to be zero, since our intuition tells us that there will always be some fluctuations for such a stochastic problem. On the other hand, if the fluctuations are extremely large, we should be suspicious!

It turns out that statisticians have calculated the probability of finding a particular value of $\chi^2$, assuming that the process involves a large number of independent random variables. The resulting probability distribution is a generalized form of the so-called normal distribution. Here the term *normal* refers to an underlying Gaussian process which generally arises in connection with random processes. In order to sketch the connection between our $\chi^2$ and the normal distribution, we

[1]Note that the present issue of quantifying how consistent the measured data are with a theoretical hypothesis is rather different from fitting the data to a theoretical formula to estimate best numerical values for the unknown parameters in the formula.

[2]These theoretical values could vary from bin to bin, but we will ignore that (largely notational) complication here.





will now digress with a bit of mathematics; if you are not interested in how the connection comes about, you may wish to skip the following section and go directly to Section G.2.

## G.1 CENTRAL LIMIT THEOREM AND THE $\chi^2$ DISTRIBUTION

A normal distribution arises in virtually all processes in which a large number of independent random values are involved. This distribution has a probability density of the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} , \tag{G.2}$$

where $f(x)dx$ is the probability that $x$ lies between $x$ and $x + dx$. It is not difficult to see that $\mu$ is the mean $\langle x \rangle$, while $\sigma^2$ is the variance or the squared fluctuation. Thus, $\sigma = \langle (x - \langle x \rangle)^2 \rangle^{1/2}$ is the standard deviation.

The fundamental mathematical result we need is the central limit theorem. We referred to this theorem briefly in Chapter 7 in connection with random processes in general. Simply put, the theorem states the following: if we have independent, identically distributed random variables $y_j$ $(j = 1, 2, 3, ..., N)$ with a finite mean $\mu_0$ and variance $\sigma_0^2$, then the distribution of its sum $x[N] \equiv \sum_{j=1}^{N} y_j$ approaches a normal distribution for large $N$, with mean $\mu = N\mu_0$ and variance $\sigma^2 = N\sigma_0^2$. Mathematically curious readers can read about the rigorous statement of this theorem as well as its proof in many statistics texts (such as Feller [1968]). Here we accept this "simple" statement without proof. The central limit theorem explains why a normal distribution is found in virtually all processes made up from a large number of constituent random processes that are independent of each other. This applies, e.g., to the random walks, molecules in a gas, and the collection of a large number of random numbers generated by a computer.

The normal distribution (G.2) is for a scalar variable $x$, i.e., in one dimension. We can generalize it to $d$ dimensions, where the random variable $\vec{r} = (x_1, x_2, ..., x_d)$ has $d$ independent components $x_1, ..., x_d$. Taking $\langle \vec{r} \rangle = (0, 0, ...)$ for simplicity, the $d$-dimensional normal distribution would be

$$f_d(\vec{r}) = \frac{1}{(2\pi\sigma^2)^{d/2}} e^{-r^2/2\sigma^2} , \tag{G.3}$$

where $\sigma^2$ is the variance of each component. It is then not difficult to imagine an extension of the central limit theorem where (G.3) is the limiting probability density (in the $d$-dimensional space) for $\vec{r}[N]$ when it is the sum of a large number $N$ of independent $d$-dimensional random vectors $\vec{y}_j$ $(j = 1, 2, ..., N)$:

$$\vec{r}[N] = \sum_{j=1}^{N} \vec{y}_j . \tag{G.4}$$

If we are interested in the probability density for the squared modulus $r[N]^2$ instead of that for $\vec{r}[N]$, we need to integrate $f_d(\vec{r})$ over all the angles in $d$-dimensions. This leads to

$$f_d(r^2) = \frac{1}{\sigma^d 2\,\Gamma(d/2)} \left(\frac{r^2}{2}\right)^{d/2-1} e^{-r^2/2\sigma^2} , \tag{G.5}$$

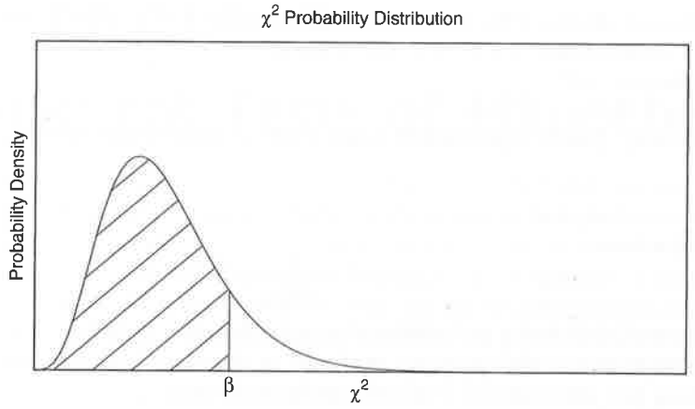


**FIGURE G.1:** The probability density of the $\chi^2$ distribution is illustrated above. The shaded area corresponds to the cumulative probability that $\chi^2$ falls somewhere between 0 and $\beta$. Sometimes you will find tables of values corresponding to its complement, i.e., the *unshaded* area which is equal to the probability that $\chi^2$ is larger than $\beta$.

in the space of $r^2 \in [0, \infty)$ where $\Gamma(x)$ is the gamma function. If $\sigma^2 = 1$, and $d = \nu$, we can rewrite this distribution in the form of

$$f_\nu(\chi^2) \;=\; \frac{1}{2\Gamma(\nu/2)} \left(\frac{\chi^2}{2}\right)^{\nu/2-1} e^{-\chi^2/2}\,, \tag{G.6}$$

where $\chi^2$ stands for $r^2$ in (G.5). This is the probability density of the $\chi^2$ distribution of $\nu$ degrees of freedom. The $\chi^2$ distribution is often quoted in terms of its cumulative values $\int_0^{\chi^2} f_\nu(x)dx$, or its complement $\int_{\chi^2}^{\infty} f_\nu(x)dx$. By a change of variables, the former can be expressed as

$$\int_0^{\chi^2} f_\nu(u)du \;=\; \frac{1}{\Gamma(\nu/2)} \int_0^{\chi^2/2} e^{-u} u^{\nu/2-1}\, du \;=\; P(\nu/2, \chi^2/2)\,, \tag{G.7}$$

where $P(a, x)$ is known as the incomplete gamma function.

If not all the $d$ components of the individual random vectors $\vec{y}_j$ are independent but there is an overall constraint among their components common to all $\vec{y}_j$, the appropriate limiting distribution is (G.6) with $\nu = d - 1$ degrees of freedom.[3] This gives the general connection between independent random processes and the $\chi^2$ distribution. In fact, the $\chi^2$ distribution of $\nu$ degrees of freedom is nothing but the normal distribution in $\nu + 1$ dimensions with an overall constraint.

Now that we understand the character of the so-called $\chi^2$ distribution, let us consider how it relates to hypothesis testing as mentioned at the beginning of this appendix. We assume that each independent measurement of a quantity produces a number in some range. We divide this range into $d$ bins and define

[3] If there are additional constraints, the number of degrees of freedom must be further reduced.



for the $j$-th measurement a vector random variable $\vec{z}_j$ of $d$-components so that its $i$-th component is 1 if the measurement falls in bin $i$ and 0 otherwise. Since these components obey the constraint that their sum is equal to 1 (because any measurement produces a number in the range of allowed values), the sum of many such variables, which constitutes many independent measurements, generally falls within the purview of the $(d-1)$-dimensional normal distribution. To be specific, we further define another vector random variable $\vec{y}_j$ whose $i$-th component is given by

$$(\vec{y}_j)_i \;\equiv\; \frac{(\vec{z}_j)_i - \langle(\vec{z}_j)_i\rangle}{\sqrt{N\langle(\vec{z}_j)_i\rangle}}\,. \tag{G.8}$$

Then the mean of each component of $\vec{y}_j$ is zero ($\langle(\vec{y}_j)_i\rangle = 0$) and its variance is $1/N$. Thus, the premise of the $\chi^2$ distribution of $d-1$ degrees of freedom applies[4] and the quantity $\chi^2 = |\sum_{j=1}^N \vec{y}_j|^2$ is distributed according to the $\chi^2$ distribution (G.6) with $\nu = d - 1$. Moreover, we can rewrite $\chi^2$ as

$$\begin{aligned}\chi^2 \;&=\; \sum_{i=1}^{d} \left[\sum_{j=1}^{N} (\vec{y}_j)_i\right]^2 \\ &=\; \sum_{i=1}^{d} \left[\frac{\left(\sum_{j=1}^N (\vec{z}_j)_i - \sum_{j=1}^N \langle(\vec{z}_j)_i\rangle\right)^2}{\sum_{j=1}^N \langle(\vec{z}_j)_i\rangle}\right]\,.\end{aligned} \tag{G.9}$$

Finally, we can identify $\sum_{j=1}^N (\vec{z}_j)_i$ to be the random variable which counts how many of the $N$ measurements fall in bin $i$ and $\sum_{j=1}^N \langle(\vec{z}_j)_i\rangle$ as its theoretical expectation. This completes our "physicist's" derivation of how the quantity $\chi^2$ defined in (G.1) obeys the $\chi^2$ distribution (G.6) with $\nu = d - 1$.

## G.2 $\chi^2$ TEST OF A HYPOTHESIS

As seen in the previous section, the value of $\chi^2$ for a set of independent measurements divided into $\nu + 1$ bins will be distributed according to the $\chi^2$ distribution of $\nu$ degrees of freedom, $f_\nu(\chi^2)$, given in (G.6) and sketched in Figure G.1. Often this probability is expressed and tabulated in terms of the cumulative value between 0 and some limit $\beta$

$$P(\nu/2, \beta/2) \;=\; \int_0^{\beta} f_\nu(\chi^2)\, d(\chi^2)\,. \tag{G.10}$$

This corresponds to the probability that the observed $\chi^2 \leq \beta$. There is unfortunately no convenient closed-form analytic expression for the function $P(a, x)$. The behavior of $P(a, x)$ for several values of $a$ is shown in Figure G.2. This probability function involves two parameters; one of them is proportional to the maximum

[4] If the expectation values have to be replaced by their estimates calculated from the sample measurements themselves, then the number of degrees of freedom is reduced further. If $r$ parameters must be estimated from the measurements to obtain the expectation values, then the appropriate number of degrees of freedom will be $d - 1 - r$.

value of $\chi^2$ ($x = \beta/2$), while the other is related to the number of degrees of freedom in the problem, $\nu$ ($a = \nu/2$).

A typical use of the $\chi^2$-distribution is in testing of the validity of a hypothesis. We predetermine a significance level $\alpha$ (0.05 or 5% is typical), and ask whether a given set of measurements reject the hypothesis at this level or not. To do this, we first look for the *critical values* of $\chi^2$ such that the probabilities of $\chi^2$ to fall below the lower critical value or above the upper critical value are each $\alpha/2$. That is, the lower critical value is the value of $\beta$ where the cumulative probability $P(\nu/2, \beta/2) = \alpha/2$, and the upper critical value is the value of $\beta$ where $P(\nu/2, \beta/2) = 1 - \alpha/2$. If the actual value of $\chi^2$ calculated from the measurements and the theoretically expected values falls outside these two limits,[5] then we must reject the hypothesis at the level of $\alpha$. Thus, if $\chi^2$ were very large or very small, we reject the original hypothesis at a small significance level $\alpha$. Of course, even if the hypothesis were true, there would still be some probability $\alpha$ that the measured $\chi^2$ would fall outside the critical values. So $\alpha$ can be considered a factor of risk that you are taking by rejecting the hypothesis. For example, for $\alpha = 0.05$ with 10 degrees of freedom, the lower critical value is about 3.247 and the upper critical value is about 20.48 according to common tables.[6] So, if your 11-bin test gives a value of $\chi^2$ less than 3.247 or greater than 20.48, you will reject the hypothesis which led to this value of $\chi^2$ at the level of $\alpha = 0.05$, taking a 5% chance of rejecting a correct theory, or with 95% confidence. If your actual value of $\chi^2$ is much further outside of the 0.05 critical values, you could push $\alpha$ to smaller values and reject the theory with even greater confidence.

The *bin test* described here is equivalent to the test of the variance often discussed in statistics texts. For such a test, $\chi^2$ is usually defined as

$$\chi^2 = \nu(s/\sigma_0)^2 , \tag{G.11}$$

where $s$ is the standard deviation measured from $\nu + 1$ samples and $\sigma_0$ is the theoretically expected standard deviation (for the entire population). Here we test the hypothesis that the standard deviation is indeed as the theory predicts by comparing this $\chi^2$ with the critical values determined similarly as before.

In Appendix F we tested random numbers created by the True Basic function `rnd` by counting how many random numbers fall in bins of width 0.1. We can now test the hypothesis that these random numbers are uniformly distributed by using the $\chi^2$ test just described. Suppose that a sample of $N = 1000$ random numbers distributed in 10 bins results in a $\chi^2$ of 11. This value falls just below the upper critical value for the significance level of $\alpha = 0.5$. Thus we reject the hypothesis at this high level of risk, or in other words, we have very little confidence in rejecting the hypothesis, based on this one sample. Actually, with larger samples we will obtain values of $\chi^2$ which are even closer to the critical value for $\alpha = 1$, making it virtually impossible to reject the hypothesis.

An even more stringent test of the significance of a hypothesis is to actually produce the entire measured distribution of $\chi^2$ from many independent set of experiments and compare the resulting distribution with the $\chi^2$-distribution. Though it is harder to come up with a quantitative measure of confidence, it would be a better test since it is based on a whole set of $\chi^2$ measurements. Returning to the test of the `rnd` function, the graph on the right in Figure G.2 shows the *distribution* of $\chi^2$ values found by performing a bin test, such as that seen in Appendix F, many times. Here we have used 11 bins, so the number of degrees of freedom is 10. We see that the most likely value of $\chi^2$ is near $x = \chi^2/2 \sim 4.5$. Since $\nu = 10$ for this case, the curve shown on the left for $P(a, x)$, with $a = \nu/2 = 5$, is appropriate. This theoretical curve has a value of 0.5 for $x = \beta/2 \sim 4.5$. Statistical theory thus predicts that in half of our measurements, that is, for half of the bin tests, we should find a value of $\chi^2$ that is smaller than $\sim 9$. This theoretical prediction is in rough agreement with the distribution of $\chi^2$, which was actually measured, and which is shown on the right in Figure G.2.

One last remark concerns the so-called $\chi^2$ fitting, where fitting is performed with a theoretical function $\hat{y}$ containing unknown parameters (see Appendix D). In general, this is done by varying the parameter values to minimize

$$\Delta_{LS} = \sum_{i=1}^{N} \frac{[y(i) - y_{fit}(i)]^2}{\sigma(i)^2} , \tag{G.12}$$

where $y(i)$ are the measurements and $\sigma(i)$ are the standard deviations of the $y$ values at $x(i)$. Then the values of $y_{fit}(i)$ with the best-fit parameters become the theoretical predictions of the model. It is easy to see that the minimum value of $\Delta_{LS}$ is precisely the quantity $\chi^2$ that we discussed in connection to the variance test, in this case, of the best-fit model. Thus, all of our results concerning the test of the validity of a theoretical model using a $\chi^2$ distribution applies to fitting

[5] This is the so-called equal-tails test. Other tests such as a one-sided one are also used.
[6] Such as in Abramowitz and Stegun in the references.



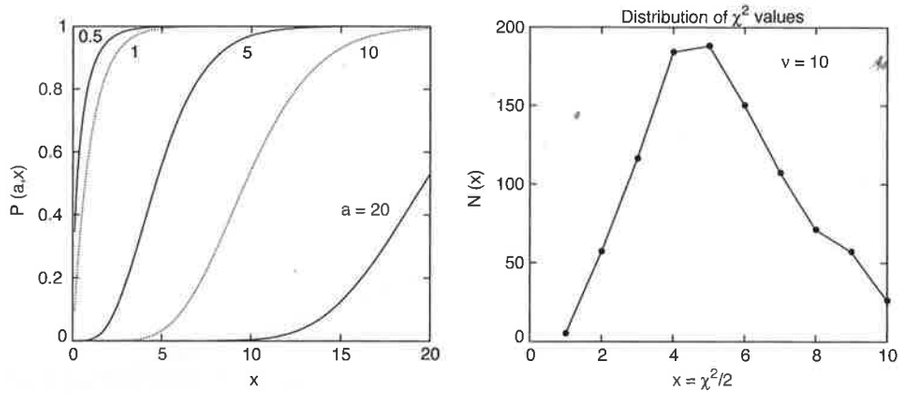


**FIGURE G.2:** Left: incomplete gamma function, $P(a, x)$. For the $\chi^2$ test, $x = \chi^2/2$ and $a = \nu/2$, where $\nu$ is the number of degrees of freedom. Right: distribution of $\chi^2$ values found by performing 1000 bin tests, as in Figure F.1. Here we used 11 bins so that $\nu = 10$ ($a = 5$), and each sequence contained 1100 numbers generated using `rnd`. This distribution was constructed in a manner similar to that employed in Figure F.3.

problems as well. It is important to note, however, that the appropriate number of degrees of freedom for the $\chi^2$ distribution is $N - n$ where $n$ is the number of parameters whose best values had to be obtained by minimizing $\Delta_{LS}$ in the first place.

## EXERCISES

**G.1.** Perform the $\chi^2$ test of your random-number generator, and compare the results with that for True Basic's `rnd` shown in the text. Is your generator better or worse? How have you made the judgment?

## REFERENCES

[1] M. Abramowitz and I. A. Stegun, Editors, 1964, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* National Bureau of Standards, Washington, DC.

[2] W. Feller, 1968, *An Introduction to Probability Theory and Its Applications, Vol. 1*, Wiley, New York. A classic introduction to the theory of probability.

[3] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1986, *Numerical Recipes*, Cambridge University Press, Cambridge. This book discusses the generation of random numbers along with numerous statistical tests.

# APPENDIX H

# Solving Linear Systems

Systems of simultaneous linear equations lie at the heart of many physical problems. The eigenvalue problems we have encountered in our studies diffusion of on fractals (Chapter 7), in quantum mechanics (Chapter 10), and in the physics of music (Chapter 11) are obvious examples, and the relaxation approach to Laplace's equation (Chapter 5) was also be cast in this form. In this Appendix we focus on two types of linear problems and briefly discuss the main methods used to attack them. For the most part, our discussions will be introductory and emphasize principles rather than give detailed how-to's (although we will, as usual, provide ample references at the end of this appendix). These computations can be quite resource hungry if the size of the problem is large, and for this reason, techniques have been developed to take advantage of almost any special features that may be present in a particular problem.[1] We will not have space to discuss most such special techniques, nor will we discuss the important issues connected with singular (or nearly singular) problems.[2]

A common type of linear problem is one where we wish to solve for a column vector $\mathbf{x}$ of $N$ components for which

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b}, \tag{H.1}$$

where $\mathbf{A}$ is an $N \times N$ matrix and $\mathbf{b}$ is a non-zero column vector also with $N$ components.[3] Another common type of linear problem is the eigenvalue-eigenfunction problem:

$$\mathbf{A} \cdot \mathbf{u}_i = \lambda_i \mathbf{u}_i, \tag{H.2}$$

where the object is to find the eigenvalues $\lambda_i$ and associated eigenfunctions $\mathbf{u}_i$ for a square $(N \times N)$ matrix $\mathbf{A}$. If $N$ linearly independent eigenvectors can be found, then we can use them to construct the inverse matrix $\mathbf{A}^{-1}$ as well. Of course these two problems are related, and some general matrix techniques are useful in both

[1] Common "special" features include tridiagonal systems, which refers to the pattern of nonzero elements in the associated matrix, and sparse systems, in which most of the elements of the matrix are zero. Sparse matrices often arise when a physical system has only local interactions.

[2] In a matrix formulation $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$, a singular problem is where the matrix $\mathbf{A}$ is singular, or its determinant is zero and it projects a set of vectors $\mathbf{x}$ into the null vector ($\mathbf{b} = \mathbf{0}$) when acting on them. A common method for dealing with singular matrices is called *singular value decomposition.*

[3] More general situations could arise where the matrix $\mathbf{A}$ is not square, i.e., it may have $M$ rows and $N$ columns where $M \neq N$, and $\mathbf{x}$ has $N$ components whereas $\mathbf{b}$ has $M$ components. This situation corresponds to the cases where the number of equations ($M$) is larger ($M > N$) or smaller ($M < N$) than the number of unknowns. Whether there will be a unique (or *any*) solution for $\mathbf{x}$ then depends on how many of the rows of $\mathbf{A}$ are linearly independent of one another and whether there are any rows that conflict. An example of the latter would be having two identical rows (say, $i$ and $j$) of $\mathbf{A}$ but with $b_i \neq b_j$.