ERRORS & STATISTICS
AY120/122 Lab notes
James R. Graham
September 2, 2008

Chapter 1

MEASURING ERRORS

Any measurement, e.g., the postion or brightness of a star is subject to uncertainty. Various physical phenomena impose fundamental limits to our measurements. The quantum nature of light means that the brightness of a given source can be established only within certain limits. The Heisenberg uncertaintly principle resticts how well we can measure positions. What is perhaps less well appreciated is that unless we quantify the uncertaintly in a measurement the value of that measurement is limited.

In this class we learn how to measure things, but just as importantly we learn how to establish how confident we are of a particular measurement. Thus, we will strive to measure the **errors** in any data that we collect. If the errors are too large for our purposes, then they must be reduced by improved techniques and repeated measurements.

Error is defined by Webster as "the difference between an observed or calculated value and the true value." Usually, we do not know the "true" value; otherwise there would be no reason for performing the experiment. Thus we must always determine in a systematic way from the data and the experimental conditions *themselves* how much confidence we can have in our experimental results.

There is one type of error that we can deal with immediately: those that originate from mistakes in measurement or computation. These errors are usually apparent either as obviously incorrect data points or as results that are not reasonably close to expected values. Rather, we are considering **random errors**, or more properly random fluctuations in our measurements, and **systematic errors**. These are two distinct types of error and they have distinctly different effects on our results: precision (for random errors) and

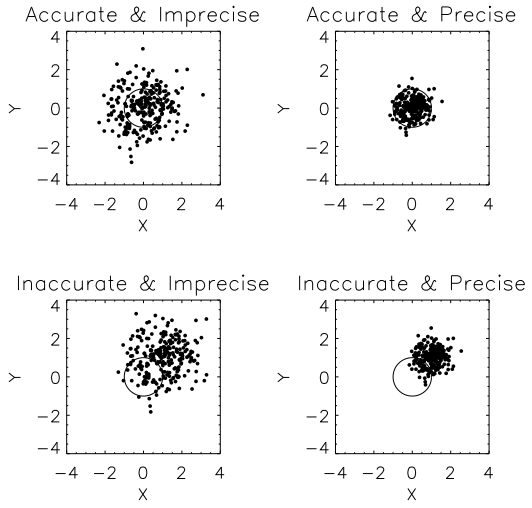

Figure 1.1: The difference between precision and accuracy is illustrated. Imagine that this example shows the distribution of arrows in a target. A precise archer places arrows in a narrow cluster. Only if the archer is also accurate will they be centered on the target.

accuracy (for systematic errors).

1.1 Accuracy and precision

It is important to distinguish between accuracy and precision. The **accuracy** of an experiment is a measure of how close the result of the experiment is to the true value. Therefore, it is a measure of the correctness of the result. The **precision** of an experiment is a measure of how well the result has been determined, without reference to its agreement with the true value. The precision is also a measure of the reproducibility of the result. The difference between these is obvious upon inspection of Figure 1.1.

It is obvious that we must consider the accuracy and precision simultaneously for any experiment. It would be a waste of time to determine a result with high precision if we knew that the result would be highly inaccurate. Conversely, a result cannot be considered to be extremely accurate if the precision is low.

1.2 Systematic errors

The accuracy of an experiment depends on the systematic errors. These may result from faulty calibration of equipment, from bias on the part of the observer, or a host of other reasons. Errors of this type are not easy to detect and not easily studied by statistical analysis. They must be estimated from an analysis of the experimental conditions and techniques. A major part of the planning of an experiment should be devoted to understanding and reducing sources of systematic errors.

Example: A student measures the length of a table top with a steel meter stick and finds that the average of her measurements yields a result of 1.982 m for the length of the table. She then learns that the meter stick was calibrated at 295 K and has an expansion coefficient of 0.0005 K^{-1}. Because her measurements were made at a room temperature of 290 K, she multiplies her results by $1+0.0005\times(290-295) = 0.9975$ so that her new determination of the length is 1.977 m. Not making this correction leads to a systematic error.

When the student repeats the experiment, she discovers another systematic error: her technique for reading the meter stick was faulty in that she did not always read the divisions from directly above the scale. By experimentation she determines that this consistently resulted in a reading that was 2 mm too short. The corrected result is 1.979 m.

In this experiment, the first result was given with a fairly high precision. The table top was found to be 1.982 m long, with a relative precision of about 1/2000, indicated by the fact that four significant figures were quoted. The corrections to this result were meant to improve the accuracy by compensating for known sources of deviation of the first result from the best estimate possible.

1.3 Random errors

The precision of an experiment is dependent on how well we can overcome random errors. These are the fluctuations in observations that yield results that differ from experiment to experiment and that require repeated experimentation to yield precise results. A given accuracy implies an equivalent precision and, therefore, also depends to some extent on random errors.

The problem of reducing random errors is essentially one of improving

the experimental method and refining the techniques, as well as simply repeating the experiment. If the random errors result for instrumental reasons, then they can be reduced by using more reliable and more precise measuring instruments. If the random errors result from statistical fluctuations associated with counting finite numbers of events, then they may be reduced by counting more events.

In the measurement of the length of the table, the student might attempt to improve the precision of her measurements by using a magnifying glass to read the scale, or she might attempt to reduce statistical fluctuations in her measurements by repeating the measurement several times. In neither case would it be useful to reduce the random errors much below the systematic errors, such as those introduced by the calibration of the meter stick. The limits imposed by systematic errors are important considerations in planning and performing experiments.

1.4 Significant figures and roundoff

The precision of an experimental result is implied by the number of digits recorded in the result, although generally the uncertainty should be quoted specifically as well. The number of significant figures in a result is defined as follows:

1. The leftmost nonzero digit is the most significant digit.

2. If there is no decimal point, the rightmost nonzero digit is the least significant digit.

3. If there is a decimal point, the rightmost digit is the least significant digit, even if it is a 0.

4. All digits between the least and most significant digits are counted as significant digits.

For example, the following numbers each have four significant digits: 1234, 123400, 123.4, 1001, 1000., 10.10, 0.0001010, 100.0. If there is no decimal point, there are ambiguities when the rightmost digit is 0. Thus, the number 1010 is considered to have only three significant digits even though the last digit might be physically significant. This can lead to ambiguity: for

example, when Sir Edmond Hillary measured the height of Mount Everest, he obtained 29,000 feet exactly—but in order to make clear that his measurement was good to 5 places he advertised 29,002. To avoid ambiguity, it is better to supply decimal points or to write such numbers in scientific notation. Thus, Hillary would have better advertised 29000. or, better, 2.9000×10^4 feet; or, even better, to renounce Imperial Units and use 8839.2 m.

When quoting an experimental result, the number of significant figures should be approximately one more than that dictated by the experimental precision. The reason for including the extra digit is to avoid errors that might be caused by rounding errors in later calculations. If the result of the table measurement $L = 1.979$ m with an uncertainty of 0.012 m, this result could be quoted as $L = (1.979 \pm 0.012)$ m. However, if the first digit of the uncertainty is large, such as 0.082 m, then we should probably quote $L = (1.98 \pm 0.08)$ m. In other words, we let the uncertainty define the precision to which we quote our result.

When insignificant digits are dropped from a number, the last digit retained should be rounded off for the best accuracy. To round off a number to fewer significant digits than were specified originally, we truncate the number as desired and treat the excess digits as a decimal fraction. Then:

1. If the fraction is greater than 1/2, increment the new least significant digit.

2. If the fraction is less than 1/2, do not increment.

3. If the fraction equals 1/2, increment the least significant digit in the rounded number only if that digit is odd.

If we did not have rule (3) then we would round down for 1, 2, 3, 4, or 4/10 th's of the time and for 5, 6, 7, 8, 9 or 5/10th's of the time we would round up. Always incrementing the least significant digit for a fracton of 1/2 or greater would lead to a systematic error. Thus according to rule (3) 1.235 and 1.245 both become 1.24 when rounded off to three significant figures, but 1.2451 becomes 1.25 because 51/100 is greater than 1/2.

1.5 The theoretical model

Above, we have concentrated on one type of uncertainty, namely those associated with the measuring process. There is one more major type, which

is associated with the intepretation and theoretical description of our result. If we don't have the correct **theoretical model**, our results may not mean much.

For example, in our description of measuring the table, we had in mind a particular theoretical model, namely a *rectangular table*. Such a table is described by two parameters, length and width, and both are well-defined. But suppose the table is a *symmetric trapezoid*—then what? Then there are two lengths, and a total of three parameters for the table: width, length of the longer long side, and length of the shorter long side.

If we measure only one length and one width, applying these results to the wrong theoretical model, we are making a major error in interpretation. This error does not show up in the quoted accuracy or precision. But it is a more serious error—an error in concept. In common parlance this is called "missing the forest for the trees". This is similar to systematic errors, which are difficult to detect and not easily studied by statistical analysis. But using the wrong theoretical model is even more difficult to discover than the presence of systematic errors, because one is making a measurement with this preconceived theoretical model in mind.

1.5.1 The Shape and Size of the Earth

An example of how a model effects interpretation of data comes from early astronomy. The first recorded measurement of the radius of the earth is by the Greek astronomer Eratosthenes (276—195 BCE). After study in Alexandria and Athens, Eratosthenes settled in Alexandria about 255 BCE and became director of the great library there. He worked out a calendar that included leap years, and he tried to fix the dates of literary and political events since the siege of Troy. Eratosthenes flourished in the century following Alexander. His writings include a poem inspired by astronomy, as well as works on the theater and on ethics. Eratosthenes was afflicted by blindness in his old age, and he is said to have committed suicide by starving himself to death.

Eratosthenes was a geographer—his maps were among the best made during the classical period. He noticed that at the town of Syene (now called Aswan), about 800 km south of Alexandria in Egypt, the Sun's rays fall vertically at noon at the summer solstice. Eratosthenes noted that at Alexandria, at the same date and time, buildings and towers cast shadows, i.e., the sun was not exactly overhead. The length of the shadow of a tower was about one eighth of this height. He correctly assumed the Sun's distance

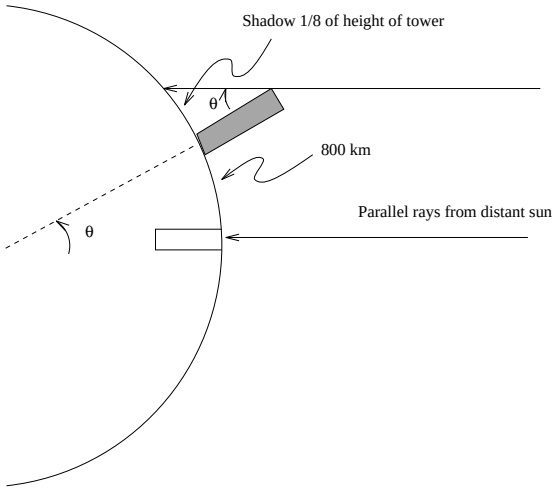

Figure 1.2: Eratosthenes' method for determining the radius of the Earth by application of similar triangles.

to be very great; its rays therefore are practically parallel when they reach the Earth. Given estimates of the distance between the two cities, he was able to calculate the circumference of the Earth. Eratosthenes measured length in units of stades, and his estimate of the distance from Alexandria to Syene was 5000 stades. The exact length of the units he used is doubtful (a Greek stadium varied in length locally from 154 to 215 m) so that the accuracy of his result is therefore uncertain because of this *systematic* error which has crept in.

Application of similar triangles is the heart of Eratosthenes' method (see Figure 1.2). However, you can see that picking the correct way to interpret the data is crucial too. Eratosthenes assumed that the Sun was very distant and that the Earth was a sphere. We call this the model—not in the sense that a model airplane is a scaled down version of the real thing, but as a set of physical assumptions that defines the system. For example, Eratosthenes could have picked quite a different model—a flat Earth and used this data to find out how far away the Sun was (see Figure 1.3). By the same type of geometric reasoning we would deduce that the sun was 8×800 km $=$ 6400 km distant. The geometry is correct, but the conclusion is wrong. This is a classic case of—garbage in garbage out!

One very important job of experimentalists and observers is to determine the parameters that describe a particular theoretical model; this is

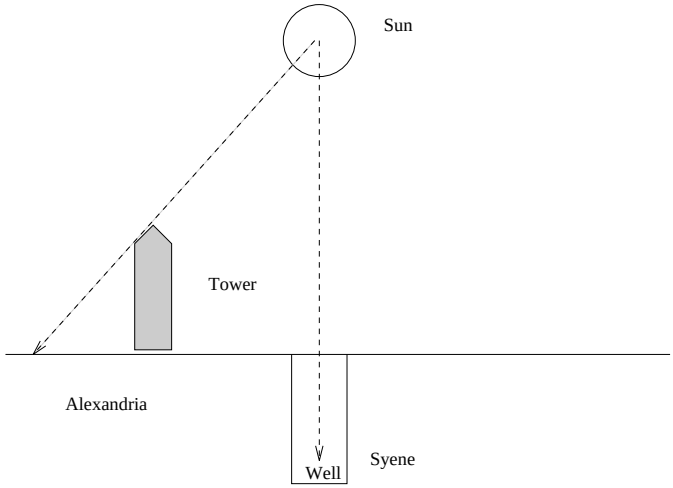

Figure 1.3: By assuming a different model Eratosthenes could have used his data to find the distance to the sun. Clearly adoption of the wrong model leads to erroneous conclusions.

how nearly all of their time and effort is spent. But they must always keep in mind that the theoretical model they have in mind may be incorrect or incomplete, and pursue this possibility experimentally. For example, in current astronomy, theorists have predicted the angular scales of fluctuations in the cosmic background radiation and nearly all observers restrict their measurements to those ranges of predicted scale—thus missing the possibility of revolutionizing the theory by discovering a major discrepancy that does not fit the theoretical model. This is in part because funding agencies are reluctant to fund "risky" experiments that have a low probability of success. These aspects are embedded in the culture of modern science and it is the rare scientist who can manage to break away from these constraints.

To summarize: our measurement uncertainties don't mean much unless we use the correct theoretical model to interpret our results.

1.6 Minimizing Uncertainties and Best Results

In virtually every case of a measurement one is attempting to determine quantitative information about some theoretical model of the object being mea-

sured. In particular, any model is quantitatively described by **parameters**. For example, astronomers are fond of measuring distances and velocities of distant galaxies to determine two basic parameters of the Big Bang model, the Hubble constant and its change with distance, called the deceleration parameter. Our student is describing the theoretical model of a rectangular table by two parameters, its length and width.

The astronomers' example is more general in the following sense. The student is measuring the two table parameters *directly*. In contrast, the astronomer is measuring the Hubble constant *indirectly* by combining observations of Doppler shift and distance, each of which has its own error. These measured quantities *and their errors* feed into the derived value of the Hubble constant *and its error*.

Thus, in general, we shall be concerned with extracting from data the best estimates and errors of parameters that describe a particular theoretical model. We shall want to understand the effect of the errors in the *data* on the errors in the *parameters*. The techniques of error analysis will help us to determine the optimum estimates of parameters to describe the data.

Chapter 2

ANALYZING AND REDUCING UNCERTAINTIES

The term error signifies a deviation of the result from some "true" value. Usually we cannot know what the true value is, and we can consider only estimates of the errors inherent in the experiment. If we repeat an experiment, the results will almost certainly differ from those of the first attempt. We can express this difference as a discrepancy between the two results. Discrepancies arise because we can determine a result only with a given uncertainty.

A study of the distribution of the results of repeated measurements of the same quantity will lead to an understanding of the uncertainties in the measurements, and the uncertainties will serve as estimates of the errors. The quoted error is thus a measure of the spread of the distribution of repeated measurements. Because, in general, we shall not be able to quote the actual error of the results, we must develop a consistent method for determining and quoting the estimated error. We must also realize that the model from which we calculate theoretical parameters to describe the results of our experiment may not be the correct model.

In the following sections we shall discuss hypothetical parameters and probable distributions of errors pertaining to the "true" state of affairs, and we shall discuss methods of making experimental estimates of these parameters and the uncertainties associated with these determinations.

2.1 Parent & Sample Distributions

If we make a measurement x_1 of a quantity x, we expect our observation to approximate the quantity, but we do not expect the experimental data point to be exactly equal to the quantity. If we make another measurement, x_2, we expect to observe a discrepancy between the two measurements because of random errors, and we do not expect either determination to be exactly correct, that is, equal to x. As we make more and more measurements, a pattern will emerge from the data. Some of the measurements will be too large, some will be too small. On the average, however, we expect them to be distributed around the correct value, assuming we can neglect or correct for systematic errors.

If we could make an infinite number of measurements, then we could describe exactly the distribution of the data points. This is not possible in practice, but we can hypothesize the existence of such a distribution that determines the probability of getting any particular observation in a single measurement. This distribution is called the **parent distribution**. Similarly, we can hypothesize that the measurements we have made are samples from the parent distribution and they form the sample distribution. In the limit of an infinite number of measurements, the sample distribution becomes the parent distribution.

Example: A student makes 100 measurements of the length of a wooden block. His observations, corrected for systematic errors, range from about 18 to 22 cm, and many of the observations are identical. Figure 2.1 shows a *histogram* or frequency plot of a possible set of such measurements. The height of each step represents the number of measurements that fall between the two values indicated by the upper and lower limits of the bar on the abscissa of the plot. If the distribution results from random errors in measurement, then it is likely that it can be described in terms of the Gaussian or normal error distribution, the familiar bell-shaped curve of statistical analysis. A Gaussian curve based on these measurements is plotted as a continuous solid line.

We identify the smooth solid curve, determined from the set of measurements displayed in the histogram, as the sample distribution. The measured data and the curve derived from them clearly do not agree exactly. The coarseness of the experimental histogram distinguishes it at once from the smooth theoretical Gaussian curve. We might imagine that, if the student were to make a great many measurements so that he could plot the his-

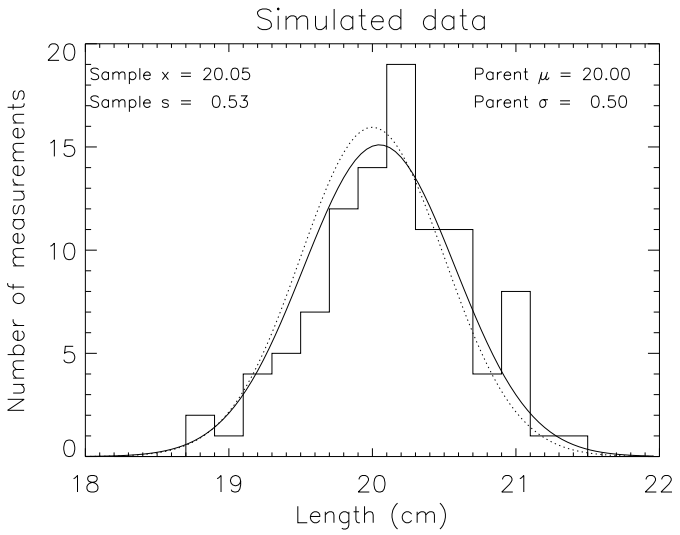

Figure 2.1: Histogram of measurements of length (steps drawn with a solid line). The smooth solid smooth curve is the continuous distribution inferred from the sample, i.e., the solid line represents a Gaussian with the same mean and standard deviation as the measurements. The dotted line represents the parent distribution from which the data are drawn.

togram in finer and finer bins, that under ideal circumstances the histogram would eventually approach a smooth Gaussian curve. If we were to calculate the parameters from such a large sample, we could determine the parent distribution, represented by the smooth dotted line in Figure 2.1.

The smooth curves in Figure 2.1 represents the probability of obtaining values of the variable x from a set of measurements. The area under the curve at the point x bounded by a range dx gives the number of events expected in that region from a 100-event sample. The area, divided by the total area of the plot, is the probability $P(x)\ dx$ that a randomly selected measurement will yield an observed value of x within the range $(x-dx/2) \leq x < (x+dx/2)$. It is convenient to think in terms of a probability density function, which, for our sample parent population, is just the dashed curve of Figure 2.1 normalized to unit area.

2.2 Notation

A number of parameters of the parent distribution have been defined by convention. We shall use Greek letters to denote them, and Latin letters to denote experimental estimations of them.

In order to determine the parameters of the parent distribution, we shall assume that the parameters of the experimental distribution equal the parameters of the parent distribution in the limit of an infinite number of measurements. If there are N observations in a given experiment, then we can denote this by

$$(\text{parent parameter}) = \lim_{N\to\infty}(\text{experimental parameter}). \tag{2.1}$$

If we make N measurements x_1, x_2, x_3, and so forth, up to a final measurement x_N, then the sum of all these measurements is

$$\sum_{i=1}^{N} x_i = x_1 + x_2 + x_3 + ... + x_N, \tag{2.2}$$

where the left-hand side is interpreted as the sum of the observations x_i over the index i from $i = 1$ to $i = N$ inclusive. Because we shall be making frequent use of the sum over N measurements, we shall simplify the notation by omitting the index whenever we are considering a sum where the index i runs from 1 to N;

$$\sum x_i \equiv \sum_{i=1}^{N} x_i \tag{2.3}$$

2.3 Mean, Median, and Mode

With these definitions, the **mean** $\bar{x}$ of the experimental distribution is given as

$$\bar{x} = \frac{1}{N}\sum x_i \tag{2.4}$$

and the mean μ of the parent population is defined as the limit

$$\mu = \lim_{N\to\infty}\left(\frac{1}{N}\sum x_i\right) \tag{2.5}$$

The mean is therefore equivalent to average value of x.

The **median** of the parent population $\mu_{1/2}$ is defined as that value for which, in the limit of an infinite number of determinations x_i, half the observations will be less than the median and half will be greater. In terms of the parent distribution, this means that the probability is 50% that any measurement x_i will be larger or smaller than the median

$$P(x_i < \mu_{1/2}) = P_i(x \geq \mu_{1/2}) = 1/2 \tag{2.6}$$

so that the median line cuts the area of the probability density distribution in half. Because of inconvenience in computation, the median is not often used as a statistical parameter.

The **mode**, or most probable value μ_{max}, of the parent population is that value for which the parent distribution has the greatest value. In any given experimental measurement, this value is the one that is most likely to be observed. In the limit of a large number of observations, this value will probably occur most often

$$P(\mu_{max}) \geq P(x \neq \mu_{max}) \tag{2.7}$$

For a symmetrical distribution these parameters would all be equal by the symmetry of their definitions. For an asymmetric distribution the median generally falls between the most probable value and the mean. The most probable value corresponds to the peak of the distribution, and the areas on either side of the median are equal.

2.4 Deviations

The **deviation** d_i of any measurement x_i from the mean μ of the parent distribution is defined as the difference between x_i and μ,

$$d_i \equiv x_i - \mu \tag{2.8}$$

For computational purposes, deviations are generally defined with respect to the mean, rather than the median or most probable value. If μ is the true value of the quantity, d_i is also the true error in x_i. The average of the deviations $\bar{d}$ must vanish by virtue of the definition of the mean

$$\lim_{N\to\infty} \bar{d} = \lim_{N\to\infty} \left[\frac{1}{N}\sum(x_i - \mu)\right] = \lim_{N\to\infty} \left(\frac{1}{N}\sum x_i\right) - \mu = 0 \tag{2.9}$$

A parameter that is easy to use analytically and that can be justified fairly well on theoretical grounds to be a more appropriate measure of the dispersion of the observations is the **standard deviation** σ. The **variance** σ^2 is defined as the limit of the average of the squares of the deviations from the mean μ,

$$\sigma^2 = \lim_{N\to\infty}\left[\frac{1}{N}\sum(x_i-\mu)^2\right] = \lim_{N\to\infty}\left(\frac{1}{N}\sum x_i^2\right) - \mu^2 \tag{2.10}$$

and the standard deviation σ is the square root of the variance. The second form is often described as the average of the squares minus the square of the average. The standard deviation is the root mean square of the deviations, and is associated with the second moment of the data about the mean. The corresponding expression for the standard deviation s of the sample population is given by

$$s^2 = \frac{1}{N-1}\sum(x_i-\bar{x})^2 \tag{2.11}$$

where the factor $N-1$, rather than N, is required in the denominator to account for the fact that the parameter x has been determined from the data and not independently.

Higher order moments involve the $\bar{d}^3$ (skewness) and the $\bar{d}^4$ (kurtosis).

2.5 Significance

The mean and the standard deviation, as well as the median, the most probable value, and the standard deviation, are all parameters that characterize the information we are seeking when we perform an experiment. Often we wish to describe our distribution in terms of just the mean and standard deviation. The mean may not be exactly equal to the datum in question if the parent distribution is not symmetrical about the mean, but it should have the same characteristics. If a more detailed description is desired, it may be useful to compute higher moments about the mean.

In general, the best we can say about the mean is that it is one of the parameters that specifies the probability distribution: It has the same units as the "true" value and, in accordance with convention, we shall consider it to be the best estimate of the "true" value under the prevailing experimental conditions.

The variance σ^2 and the standard deviation σ characterize the uncertainties associated with our experimental attempts to determine the "true"

values. For a given number of observations, the uncertainty in determining the mean of the parent distribution is proportional to the standard deviation of that distribution. The standard deviation σ is, therefore, an appropriate measure of the uncertainty due to fluctuations in the observations in our attempt to determine the "true" value. But these two parameters are not the full story.

Chapter 3

THE PROBABILITY FUNCTION

In general, the distribution resulting from *purely statistical errors* can be described well by the two parameters, the mean and the standard deviation. However, we should be aware that, at distances of a few standard deviations from the mean of an experimental distribution, *nonstatistical errors* may dominate. In some cases, it may be preferable to describe the spread of the distribution in terms of the mean absolute deviation, rather than the standard deviation, because the former tends to deemphasize measurements that are far from the mean. There are also distributions for which the variance does not exist. The mean absolute deviation or some other quantity must be used as a parameter to indicate the spread of the distribution in such cases.

We describe the distribution of all errors, both statistical and nonstatistical, by the probability function $P(x)$ of the parent population. The mean μ and the standard deviation σ are simple moments of this function. The probability function $P(x)$ is defined such that in the limit of a very large number of observations, the fraction dN of observations of the variable x that yield values between x and $x + dx$ is given by $dN = P(x)dx$.

The mean μ is the **expectation value** $\langle x \rangle$ of x, and the variance σ^2 is the expectation value $\langle (x - \mu)^2 \rangle$ of the square of the deviations of x from μ. The expectation value $\langle f(x) \rangle$ of any function of x is defined as the weighted average of $f(x)$, over all possible values of the variable x, with each value of $f(x)$ weighted by the probability density distribution $P(x)$.

3.1 Continuous Distributions

If the probability function is a continuous smoothly varying function $P(x)$ of the observed value x, then the mean μ is the **first moment** of the parent distribution

$$\mu = \int_{-\infty}^{\infty} xP(x)dx, \tag{3.1}$$

and the variance σ^2 is the **second moment**

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 P(x)dx = \int_{-\infty}^{\infty} x^2 P(x)dx - \mu^2. \tag{3.2}$$

The expectation value of any function of x is

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x)P(x)dx. \tag{3.3}$$

3.2 Discrete Distributions

Our measurements are always discrete. In this case, we consider the different measured values to lie in bins. That is, if a measured value is x, we assign it to a bin numbered j; bin j has boundaries $x_j \pm \delta x_j/2$, where x_j is the central value and δx_j is the width of the bin. We define n bins, with bin $j = 0$ containing the minimum measured value and bin $j = n-1$ containing the maximum. Then we replace the integrals by sums.

Suppose there are N observed values of x. Then we should expect each bin to contain $NP(x_j)$ of the observations. That is,

$$\mu = \lim_{N\to\infty} \sum x_i = \lim_{N\to\infty} \frac{1}{N} \sum \left[x_j N P(x_j)\right] = \lim_{N\to\infty} \sum \left[x_j P(x_j)\right]. \tag{3.4}$$

Similarly, the variance σ^2 can be expressed in terms of the probability function $P(x)$,

$$\sigma^2 = \sum \left[(x_j - \mu)^2 P(x_j)\right] = \sum \left[x_j P(x_j)\right]^2 - \mu^2. \tag{3.5}$$

In general, the expectation value of any function f(x) of x is given by

$$\langle f(x) \rangle = \sum \left[f(x_j)P(x_j)\right] \tag{3.6}$$

3.3 Observed Distribution vs. Parent-Population Distribution

What is the connection between the probability distribution of the parent population and an experimental sample we obtain? We have already seen that the uncertainties of the experimental conditions preclude a determination of the "true" values themselves. As a matter of fact, there are three levels of abstraction between the data and the information we seek:

1. The *observed* probability distribution. From our experimental data points we can determine a sample frequency distribution that describes the way in which these particular data points are distributed over the range of possible data points. We use $\bar{x}$ to denote the mean of the data and s^2 to denote the sample variance. The shape and magnitude of the sample distribution vary from sample to sample.

2. The *parent* probability distribution. From the parameters of the sample probability distribution we can estimate the parameters of the probability distribution of the parent population of possible observations. Our best estimate for the mean μ is the mean of the sample distribution $\bar{x}$, and the best estimate for the variance σ^2 is the compensated variance s^2. The shape of this parent distribution must be estimated or assumed.

3. The *parameters* of the theoretical model. From the estimated parameters of the parent distribution we estimate the results sought—the parameters of the theoretical model.

Refer to Figure 2.1 which shows a histogram of the measurements of the length of the block and two Gaussian curves. The calculation of the smooth solid curve was based on the parameters $x = 20.05$ cm and $s = 0.53$ cm determined experimentally from the data displayed in the histogram. The dashed curve was based on the parameters $\mu = 20.00$ cm and $\sigma = 0.50$ cm of the parent distribution, which could have been obtained by making a very large number of measurements. The difference between the experimental mean $\bar{x}$ and the "true" mean μ is obvious on comparing the two curves.

We should note, however, that even the definition of μ might be somewhat uncertain, because the edge of the block is not perfectly smooth and

we should have to define exactly what we mean by the length. In other words, even if the theoretical model of a rectangular table is correct, this model does not include the roughness of the edge. This makes the observed probability distribution differ from the theoretical one. We are always restricted, in working with experimental data, to estimating the parameters of the parent population, and sometimes to defining what we mean by the parent population.

Nevertheless, by considering the data to be a sample from the parent population, we can estimate the shape and dispersion of the parent distribution to obtain useful information on the precision and reliability of our results. Thus, we find the sample mean $\bar{x}$ as an estimate of the mean μ in order to find the "true" value of length of the block, and we find the sample variance s^2 as an estimate of the variance σ^2 in order to estimate the uncertainty in our value for μ.

3.4 Probability Distributions

Of the probability distributions that are involved in the analysis of experimental data, the **binomial distribution**, the **Poisson distribution**, and the **Gaussian distribution** are the most important. The Gaussian or normal error distribution is the most common because it describes the distribution of random observations for many experiments, as well as describing the distributions obtained when we try to estimate the parameters of most other probability distributions.

The Poisson distribution is appropriate for counting experiments where the data represent the number of items or events observed per unit interval. It is important in the study of random processes such as the arrival of photons. It is therefore of great interest to astronomers.

The binomial distribution is applied to experiments in which the result is one of a small number of possible final states, such as the number of "heads" or "tails" in a series of coin tosses. Because both the Poisson and the Gaussian distributions can be considered as limiting cases of the binomial distribution, we shall derive the binomial distribution.

3.4.1 BINOMIAL DISTRIBUTION

Suppose we toss a coin in the air. There is a 50% probability that it will land heads up and a 50% probability that it will land tails up. By this we mean that if we toss a coin repeatedly, the fraction of times that it lands with heads up will asymptotically approach 1/2. For any given toss, the probability does not determine whether or not it will land heads up; it can only describe how we should expect a large number of tosses to be divided into two possibilities.

Suppose we toss two coins. There are now four different possible permutations of the way in which they can land: both heads up, both tails up, and two mixtures of heads and tails depending on which one is heads up. Because each of these permutations is equally probable, the probability for any choice of them is 1/4. To find the probability for obtaining a particular mixture of heads and tails, without differentiating between the two kinds of mixtures, we must add the probabilities corresponding to each possible kind. Thus, the total probability of finding either head up and the other tail up is 1/2. Note that the sum of the probabilities for all possibilities (1/4 + 1/4 + 1/4 + 1/4) is always equal to 1 because something is bound to happen.

Suppose we toss n coins, where n is some integer. What is the probability that exactly x coins will land heads up, without distinguishing which of the coins actually belongs to which group? We can consider the probability $P(x;n)$ to be a function of the number n of coins tossed and of the number x of coins that land heads up. For a given experiment in which n coins are tossed, this probability $P(x;n)$ will vary as a function of x.

3.4.2 Permutations and Combinations

If n coins are tossed, there are 2^n different possible ways in which they can land. This is because the first coin has two possible orientations, for each of these the second coin also has two such orientations, for each of these the third coin also has two, and so on. Because each of these possibilities is equally probable, the probability for any one of these possibilities to occur at any toss of n coins is $1/2^n$.

How many of these possibilities will contribute to our observations of x coins with heads up? Imagine two boxes, one labelled "heads" and divided into x slots, and the other labelled "tails". We shall consider first the question of how many **permutations** of the coins result in the proper separation of

x in one box and $n - x$ in the other.

In order to count the number of permutations $Pm(n, x)$, pick up the coins one at a time from the n coins and put x of them into the "heads" box. We have a choice of n coins for the first one we pick up. For our second selection we can choose from the remaining $n - 1$ coins. The range of choice is diminished until the last selection of the xth coin can be made from only $n - x + 1$ remaining coins. The total number of choices for coins to fill the x slots in the "heads" box is the product of the numbers of individual choices:

$$Pm(n, x) = n(n - 1)(n - 2)...(n - x + 2)(n - x + 1) \tag{3.7}$$

This can be expressed more easily in terms of factorials

$$Pm(n, x) = \frac{n!}{(n - x)!} \tag{3.8}$$

We have calculated the number of permutations $Pm(n, x)$ that will yield x coins in the "heads" box and $n-x$ coins in the "tails" box, with the provision that we have identified which coin was placed in the "heads" box first, which was placed in second, and so on. That is, we have ordered the x coins in the "heads" box. In our computation of 2^n different possible permutations of the n coins, we are only interested in which coins landed heads up or heads down, not which landed first. Therefore, we must consider contributions different only if there are different coins in the two boxes, not if the x coins within the "heads" box are permuted into different time orderings.

The number of different **combinations** $C(n, x)$ of the permutations in the preceding enumeration results from combining the $x!$ different ways in which x coins in the "heads" box can be permuted within the box. For every $x!$ permutations, there will be only one new combination. Thus, the number of different combinations $C(n, x)$ is the number of permutations $Pm(n, x)$ divided by the degeneracy factor $x!$ of the permutations:

$$C(n, x) = \frac{Pm(n, x)}{x!} = \frac{n!}{x!(n - x)!} = \binom{n}{x} \tag{3.9}$$

This is the number of different possible combinations of n items taken x at a time, commonly referred to as $\binom{n}{x}$.

3.4.3 Probability

The probability $P(x;n)$ that we should observe x coins with heads up and $n-x$ with tails up is the product of the number of different combinations $C(n,x)$ that contribute to that set of observations multiplied by the probability for each of the combinations to occur, which we have found to be $(1/2)^n$.

Actually, we should separate the probability for each combination into two parts: one part is the probability $(1/2)^x$ for x coins to be heads up; the other part is the probability $(1/2)^{n-x}$ for the other $n-x$ coins to be tails up. The product of these two parts is the probability of the combination. In the general case (i.e., lopsided but identical coins), the probability p of success for each item (in this case landing heads up) is not equal to the probability $q = 1 - p$ for failure (landing tails up). The probability for each of the combinations of x coins heads up and $n - x$ coins tails up is $p^x q^{n-x}$.

With these definitions of p and q, the probability $P_B(x;n,p)$ for observing x of the n items to be in the state with probability p is given by the **binomial distribution**

$$P_B(x;n,p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \tag{3.10}$$

where $q = 1-p$. The name for the binomial distribution comes from the fact that the coefficients $P_B(x;n,p)$ are closely related to the binomial theorem for the expansion of a power of a sum. According to the binomial theorem,

$$(p+q)^n = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} \tag{3.11}$$

The $(j+1)$th term, corresponding to $x = j$, of this expansion, therefore, is equal to the probability $P_B(j;n,p)$. We can use this result to show that the binomial distribution coefficients $P_B(x;n,p)$ are normalized to a sum of 1. The right-hand side of this equation is the sum of probabilities over all possible values of x from 0 to n and the left-hand side is just $1^n = 1$.

3.4.4 Mean and Standard Deviation

The mean of the binomial distribution is evaluated by combining the definition of μ with the formula for the probability function

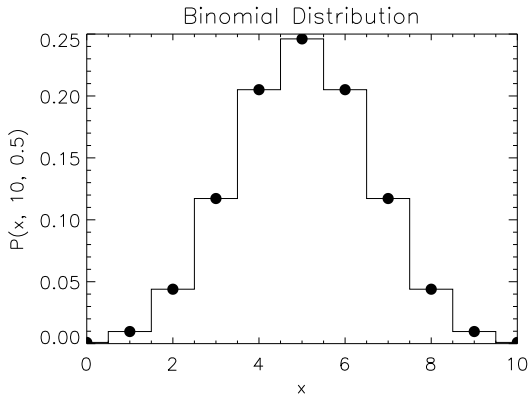

Figure 3.1: Binomial distribution for $n = 10$ and $p = 0.5$. Hence, $\mu = 5.0$ and $\sigma = 1.6$.

$$\mu = \sum_{x=0}^{n} \left[x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \right] = np \tag{3.12}$$

If we perform an experiment with n items and observe the number x of successes, after a large number of repeated experiments the average $\bar{x}$ of the number of successes will approach a mean value μ given by the probability for success of each item p times the number of items n. In the case of coin tossing where $p = 1/2$, we should expect on the average to observe half the coins land heads up, which seems eminently reasonable.

The variance σ^2 of a binomial distribution is similarly evaluated

$$\sigma^2 = \sum_{x=0}^{n} \left[(x-\mu)^2 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \right] = np(1-p) \tag{3.13}$$

If the probability for a single success p is equal to the probability for failure $p = q = 1/2$, then the distribution is symmetric about the mean μ, and the median $\mu_{1/2}$ and the most probable value are both equal to the mean. In this case, the variance σ^2 is equal to half the mean $\sigma^2 = \mu/2$. If p and q are not equal, the distribution is asymmetric with a smaller variance.

The parent distribution $P_B(x; 10, 1/2)$ is shown in Figure 3.1. The curve is symmetric about its peak at the mean and the magnitudes of the points are such that the sum of the probabilities over all the points is equal to 1.

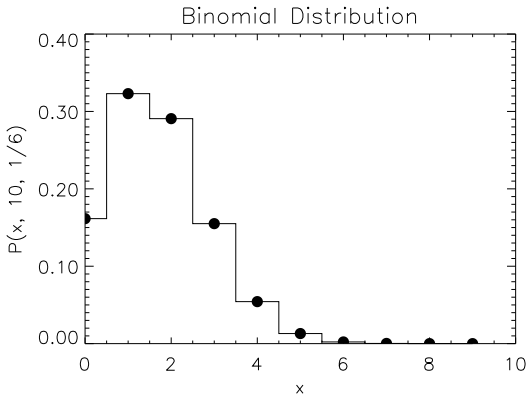

Figure 3.2: Binomial distribution for $n = 10$ and $p = 1/6$. Hence $\mu = 10/6 = 1.67$ and $\sigma = 1.2$.

The mean μ is:

$$\mu = np = 10 \times (1/2) = 5, \tag{3.14}$$

and the standard deviation σ is

$$\sigma = \sqrt{np(1-p)} = \sqrt{10 \times 0.5 \times 0.5} \simeq 1.58. \tag{3.15}$$

Suppose we roll 10 dice. What is the probability that x of these dice will land with the 1 up? If we throw one die, the probability of its landing with 1 up is $p = 1/6$. If we throw 10 dice, the probability for x of them landing with 1 up is given by the binomial distribution $P_B(x; n, p)$ with $n = 10$ and $p = 1/6$: This distribution is illustrated in Figure 3.2. The distribution is not symmetric about the mean or about any other point. The most probable value is $x = 1$, but the peak of the smooth curve occurs for a slightly larger value of x. The mean and standard deviation are

$$\mu = 10 \times 1/6 \simeq 1.67, \tag{3.16}$$

and

$$\sigma = \sqrt{10 \times (1/6) \times (5/6)} \simeq 1.18. \tag{3.17}$$

3.4.5 Poisson Distribution

The Poisson distribution represents an approximation to the binomial distribution for the special case where the average number of successes is much smaller than the possible number; that is, when $\mu \ll n$ because $p \ll 1$. For such experiments the binomial distribution correctly describes the probability $P_B(x; n, p)$ of observing x events per time interval out of n possible events, each of which has a probability p of occurring, but the large number n of possible events makes exact evaluation from the binomial distribution impossible. Furthermore, in these experiments, neither the number n of possible events nor the probability p for each is usually known. What may be known instead is the average number of events μ expected in each time interval or its estimate $\bar{x}$. The Poisson distribution provides an analytical form appropriate to such investigations that describes the probability distribution in terms of just the variable x and μ.

Consider the binomial distribution in the limiting case of $p \ll 1$. We are interested in its behavior as n becomes large while the mean $\mu = np$ remains constant.

The binomial probability function may be written as

$$P_B(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{-x} (1-p)^n \quad (3.18)$$

Recall, that the second term is

$$\frac{n!}{(n-x)!} = n(n-1)(n-2)...(n-x-1), \quad (3.19)$$

which we can consider it to be the product of x terms, each of which is very nearly equal to n because $x \ll n$ in the region of interest. This term asymptotically approaches n^x. The product of the second and third terms thus becomes $(np)^x = \mu^x$. For small p, the fourth term is approximately equal to $(1 + px)$, which tends to 1 as $p \to 0$.

The last term can be rearranged by substituting $n = \mu/p$ to show that it asymptotically approaches $e^{-\mu}$,

$$\lim_{p \to 0} (1-p)^n = \lim_{p \to 0} \left[(1-p)^{1/p} \right]^\mu = \left(\frac{1}{e} \right)^\mu = e^{-\mu} \quad (3.20)$$

Combining these approximations, we find that the binomial distribution probability function $P_B(x; n, p)$ asymptotically approaches the Poisson distribution $P_p(x; \mu)$ as $p \to 0$,

$$P_p(x;\mu) = \frac{\mu^x}{x!}e^{-\mu} \tag{3.21}$$

Because this distribution is an approximation to the binomial distribution for $p \ll 1$, the distribution is asymmetric about its mean μ and will resemble that of Fig. 3.2. Note that $P_P(x;\mu)$ does not become 0 for $x = 0$ and is not defined for negative x .

3.4.6 Mean and Standard Deviation

The Poisson distribution is a discrete distribution; it is defined only at integral values of x, although the parameter μ is a positive, real number. The mean of the Poisson distribution is actually the parameter μ that appears in the probability function $P_P(x;\mu)$. To verify this, we can evaluate the expectation value expectation value $\langle x\rangle$ of x

$$\langle x\rangle = \sum_{x=0}^{\infty}\left(x\frac{\mu^x}{x!}e^{-\mu}\right) = \mu e^{-\mu}\sum_{x=1}^{\infty}\frac{\mu^{x-1}}{(x-1)!} = \mu e^{-\mu}\sum_{y=0}^{\infty}\frac{\mu^y}{y!} = \mu \tag{3.22}$$

To find the standard deviation σ, the expectation value of the square of the deviations can be evaluated :

$$\sigma^2 = \langle(x-\mu)^2\rangle = \sum_{x=0}^{\infty}\left[(x-\mu)^2\frac{\mu^x}{x!}e^{-\mu}\right] = \mu \tag{3.23}$$

Thus, the standard deviation σ is equal to the square root of the mean μ and the Poisson distribution has only a single parameter, μ.

Chapter 4

AN EXAMPLE

The following example is designed to illuminate some aspects of the statistical nature of data, the concepts of a probability distribution function, and the mean and standard deviation.

A lighthouse emits flashes of light in a narrow beam. Suppose that the flashes occurs at random intervals as the lantern rotates so that it is equally likely that the flash occurs at any angle, θ. The probability distribution for θ is a uniform or flat distribution because all angles are equally probably. If we were to draw a histogram for θ for a large number of flashes, all bins between -90° and $+90^\circ$ would have equal numbers of events.

If we measure θ for a finite number of flashes we do not expect the histogram to be perfectly flat because of the inherent randomness of the flashes (Fig. 4.2). In this example a total of 1024 flashes were observed. Since the flashes occurs at random, all angles are equally probable and the histogram should be flat. However, in any finite experiment, there will be fluctuations about the expected average value (indicated by the horizontal dashed line). Compare the left and right hand panels of Fig. 4.2. The left hand reprents an experiment with 1024 measurements, and the right represents $1024 \times 1024 =$ 1,048,576 measurements. In this example with more than a million measurements the measured histogram tends to the theoretically expected uniform or flat distribution.

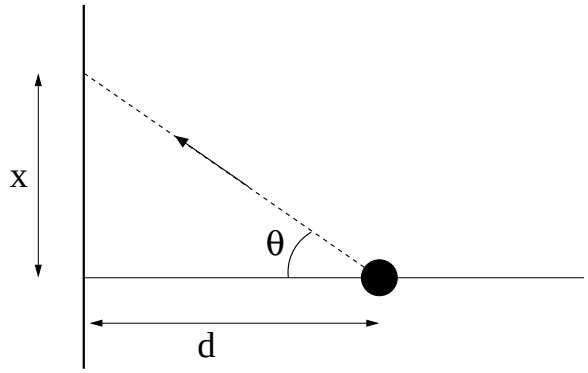

Figure 4.1: A lighthouse, represented by the black dot, emits a flash of light that is confined to a narrow beam. The light is directed towards the coast (to the left), i.e., $-90° \leq \theta \leq 90°$. The lighthouse is a distance, d, out to sea. If the pulse of light is emitted at an angle θ, then the light is detected at a position $x = d \tan \theta$.

4.1 From Histograms to Probability

A histogram is a graphical device for illustrating the relative occurrence of the outcomes of an experiment. To construct a histogram the event bins must be defined. The data can then be sorted into the bins and the number of events in each bin counted. The height of the bars in Fig. 4.2 is proportional to the number of events in that bin. Bins must be mutually exclusive, e.g., for a coin flipping experiment there are two bins corresponding to heads or tails. In some cases the bin definitions are self-evident, e.g., if you decide you would want to know what month to take your vacation you might make a histogram of the number of hours of sunshine per month. In this example an hour of sunshine is an "event" and the bins are drawn from the set of months of the year.

In other cases defining the bins can be arbitrary. This is clearly true for continuous variables, e.g., length or mass, or, as in the above example, an angle. For example, in constructing a histogram of the heights of enrollees in Astro-122 you chould choose cm or inches; since height is a continuous variable, bins of fractional unit width, e.g, 1/2 an inch, are also perfectly valid. If we pick bins that are too narrow, the resultant histogram will not provide useful visualization of the distribution of data. If bins are too narrow then the bins will contain either 1 or 0 events. Conversely, if bins are too

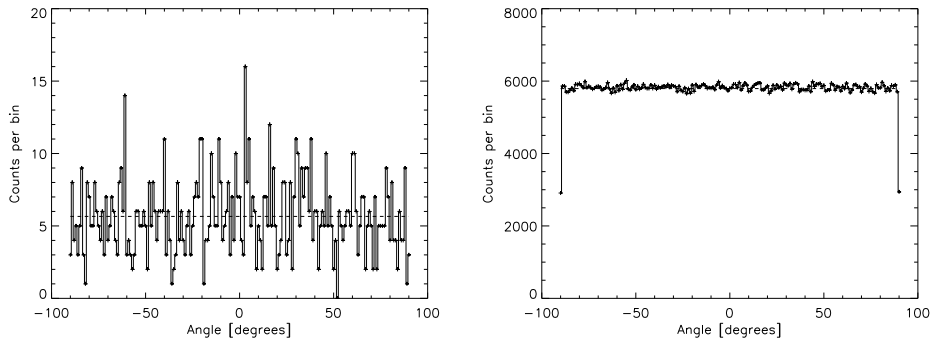

Figure 4.2: **Left**: Histogram of the angle, θ, at which narrow pulses of light are emitted from the lighthouse in Fig. 4.1. Each bin is 1 degree wide. In this example a total of 1024 flashes were observed. Since the flash occurs at random, all angle are equally probable and the histogram should be flat. However, in any finite experiment, there will be fluctuations about the expected average values (indicated by the horizontal dashed line). **Right**: Identical plot but for 1,048,576 flashes (2^{20}). As the size of the experiment grows and more measurements are made the results tend to the theoretical expectation of a flat histogram.

wide then the bulk of the data will fall into only a few bins.

Histograms are instructive because they tell us about the relative occurence of different events. If all events are equally likely, then the histogram is flat (Fig. 4.2). If a histogram is sharply peaked that means the events at the peak are more likely to occur. A histogram of rainfall by month For northern California is sharply peaked in late Winter and early Spring. That means that rain is more likely in January than in July. Thus, the height of a bar in a histogram is proportional to the **probability** of that event happening.

We introduce a further degree of abstraction by supposing that the events are drawn from a **probability distribution function**. In any finite experiment we can only estimate the probability of a particular event occuring and we cannot deduce the full form of the probability distribution function. For example, with a finite number of flips we can never be completely confident that a particular coin is fair.

Figure 4.2 represents a histogram, where the y-axis shows the counts per bin. Naturally, the scale on the y-axis changes when the number of measure-

ments changes—compare the left and right panels of Figure 4.2. An alternate way of describing the results is to cast them in terms of probability. If we use probabilities then, the y-axis scaling should remain the same, independent of how many measurements we make. Moreover, if we convert the number of counts per bin into the probabilty of getting that count then we can compare directly with a theoretical probability distribution function.

Probability can be defined in an intuitive way. Consider an experiment where there are only two outcomes, e.g., heads or tails from flipping a coin. If the coin is flipped N times then we observe N_H heads and N_T tails. The probability of getting a head is the fraction of outcomes that corresponds to that event, i.e., $p(H) = N_H/N$ and the probabiliy of getting a tail is $p(T) = N_T/N$. This cannot be exact because there are always statistical fluctuations, and we can only know the true probabilty after an infinte number of trials. The probabilities are therefore expressed as limits. The probability of getting a head $p(H)$ is

$$p(H) = \lim_{N\to\infty} \frac{N_H}{N} \tag{4.1}$$

and the probability of getting a tail $p(T)$ is

$$p(T) = \lim_{N\to\infty} \frac{N_T}{N}. \tag{4.2}$$

Note that $N_T + N_H = N$, thus $p(T) + p(H) = 1$. Not only are probabilites positive, definite quantities ($0 \geq p \geq 1$) the probability distribution fucntion must be normalized. For a discrete distrbibution, where the outcomes are labeled $i = 0...N-1$ we have

$$\sum_{i=0}^{N-1} p_i = 1. \tag{4.3}$$

This is just a way of saying that if our probability distribution function p_i is a complete description of all possible outcomes, then something must happen! Thus $p(H) = 1 - p(T)$, and if our coin is fair, then $p(H) = p(T) = 1/2$.

Returning to the lighthouse we can replot the y-axis as a probability—we simply divide the number of counts per bin by the total number of measurements made in that experiment. Be sure to note that the probability refers not just to the probability of getting so many counts at such an angle, the only probability we can compute is the probability of getting witin the range

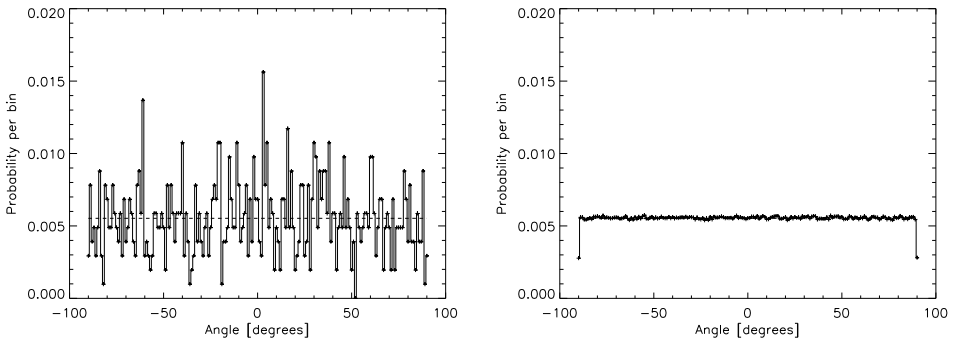

Figure 4.3: The data from Fig. 4.2 expressed as an observed probabilty. Notice that the y-axis scales are now the same and the plots are easy to compare directly. **Left**: Observed probability of the angle, θ in 1-degree wide bin. A total of 1024 flashes were observed. **Right**: The same plot but for 1,048,576 flashes (2^{20}). As the size of the experiment grows and more measurements are made the results tend to the theoretical expectation of uniform probability ($p_i = 1/180 \simeq 0.0556$).

of the bin

$$p_i = p(\theta_i - \Delta\theta/2 \leq \theta < \theta_i + \Delta\theta/2) = \lim_{N\to\infty} \frac{N(\theta_i - \Delta\theta/2 \leq \theta < \theta_i + \Delta\theta/2)}{N} \tag{4.4}$$

where θ_i denotes the center of bin i, and $\Delta\theta$ is the width of the bin.

4.1.1 Mean & Standard Deviation

A common measure of the typical value determined from a set of experimental data is the **average**. The average value of a data set θ_i, comprised of N measurements, where $i \in \{0, 1, 2, \ldots, N-1\}$, is computed as

$$X(\theta) = \frac{1}{N} \sum_{i=0}^{i=N-1} \theta_i. \tag{4.5}$$

The reason for thinking about probability distribution functions now becomes evident. If we know the form of the probability distribution function we can predict what value of the average to expect. This value is often refered to as the **mean** or **eXpectation value**. If all values between $-90°$ and

$+90^\circ$ are equally probably then the mean value must be 0 degrees. If p_j is the probability of achieving a result x_j from a discrete list of outcomes $j \in \{0, 1, 2, \ldots, M-1\}$ then the mean or expectation value is

$$\mu(x) = \sum_{j=0}^{M-1} p_j x_j. \tag{4.6}$$

Do not confuse M and N. In the previous example i was the experiment counter—every time we made a new measurement the value was labeled with i, and i was incremented by 1. Here, j enumerates the possible outcomes. In the coin tossing example, $j = 0$ refers to heads and $j = 1$ refers to tails and $M = 2$. In the lighthouse example, it is evident that the results are not natrually restricted to a finite list of 181 outcomes. Out choice of 1-degree wide bins is arbitrary, and the probability distribution funciton is a continuous function of θ. To describe the lighthouse case we should use

$$\mu(\theta) = \int_{\theta=-90^\circ}^{\theta=90^\circ} p(\theta)\theta \; d\theta, \tag{4.7}$$

or more generally the mean is

$$\mu(x) = \int_{-\infty}^{\infty} p(x)x \; dx. \tag{4.8}$$

Note that the the Greek letter μ is frequently reserved to denote the mean of the probability distribution and that X is reserved to denote the mean of a finite set of observations. The observed value X is an estimate of μ and only equals the true value μ in the limit as $N \to \infty$.

The spread of measurements is usually quantified using the **variance**, s^2. The square root of the variance is known as the **standard deviation**. The standard deviation is useful, although perhaps not as fundamental a quantity as the variance, because it has the same units as the mean. The variance of a set of N measurements x_i where $i \in 0, 1, 2, \ldots, N-1$ is

$$s^2 = X(x^2) - X(x)^2. \tag{4.9}$$

In words the varianace is the difference between the mean value of x^2 and the square of the mean value of x. The variance is also equal to the mean square deviation:

$$s^2 \;=\; \frac{1}{N} \sum_{i=0}^{i=N-1} [x_i - X(x)]^2 \tag{4.10}$$

$$\begin{aligned} &= \frac{1}{N}\sum\left[x_i^2 - 2X(x)x_i + X(x)^2\right] \\ &= \frac{1}{N}\sum x_i^2 - \frac{2X}{N}\sum x_i + X(x)^2 \\ &= X(x^2) - X(x)^2 \end{aligned} \tag{4.11}$$

In this derivation we have used the fact that X is a constant and $\sum X = NX$. For clarity we have dropped the limits of the sumation, but all sums are over the range $i = 0, N-1$.

For a continuous probability distributions it should be evident from the defintion of the variance, Eq. 4.9 that

$$\begin{aligned} \sigma^2 &= \int p(x)x^2\ dx - \left(\int p(x)x\ dx\right)^2 \\ &= \int p(x)(x-\mu)^2\ dx \end{aligned} \tag{4.12}$$

Here the limits of integration are from $-\infty$ to ∞, but again for clarity these have been dropped. For this defintion it is evident that the variance is the expectation value of the deviation from the mean, squared. Note again that the roman s^2 is distinct from the Greek μ: s^2 approaches σ^2 in the limit of infinite N.

4.1.2 The Variance of a Uniform Distribution

The mean value of a uniform distribution is easy to compute. For the example in Fig. 4.2 it is evident that $\mu(\theta) = 0$. This result would be true for any probability distribution function that is symmetric about the origin. The variance of a uniform distribution is not so obvious. First consider the simple uniform distribution

$$\begin{aligned} p(x) &= 0;\ x < 1/2, \\ &= 1;\ 1/2 \le x \le 1/2, \\ &= 0;\ x > 1/2. \end{aligned} \tag{4.13}$$

In which case it is evident that $\mu = 0$ and

$$\sigma^2 = \int_{-1/2}^{1/2} x^2 dx$$

$$= \frac{1}{3}x^3\Big|_{-1/2}^{1/2}$$
$$= \frac{1}{12}.$$

By scaling the distribution Eq. (4.13) to the lighthouse case we expect that the standard deviation is $\sigma = 180/\sqrt{12} \simeq 52.0$ degrees.

4.1.3 Transforming Probability Distributions

Although the underlying probability distribution in the lighthouse example is the simplest possible the observed quantity is not the angle θ at which the narrow pulse of light is emitted, but the location on the shore where the beam is detected, x (see Fig. 4.1). Obeys a very different probability distribution. The relation between θ and x is

$$x = d \tan\theta. \tag{4.14}$$

Consider the probability that the beam is emitted at an angle between $\theta - \delta\theta/2$ and $\theta + \delta\theta/2$. This angular range defines a bin in a histogram. The correspodnding position bin is ranges between $x - \delta x/2$ and $x + \delta x/2$. The number of events in these two correspnding bins must be the same, i.e.,

$$N_{flash}\ p(\theta)\ d\theta = N_{flash}\ p(x)\ dx, \tag{4.15}$$

which we can rearagne to show that

$$p(x) = p(\theta)\frac{d\theta}{dx}. \tag{4.16}$$

Differentiating Eq. (4.14) while holding d constant yields

$$\frac{d\theta}{dx} = \frac{d}{d^2 + x^2}, \tag{4.17}$$

and $p(\theta) = 1/\pi$ (note that we have to use radians because angles in Eq. (4.14) must be expressed in radians) we find that the probability distribution function for $p(x)$ is

$$p(x) = \frac{1}{\pi}\frac{d}{d^2 + x^2}. \tag{4.18}$$

Figure 4.4 takes the angles histogramed in Fig. 4.2 and converts them to positions using Eq. (4.14).

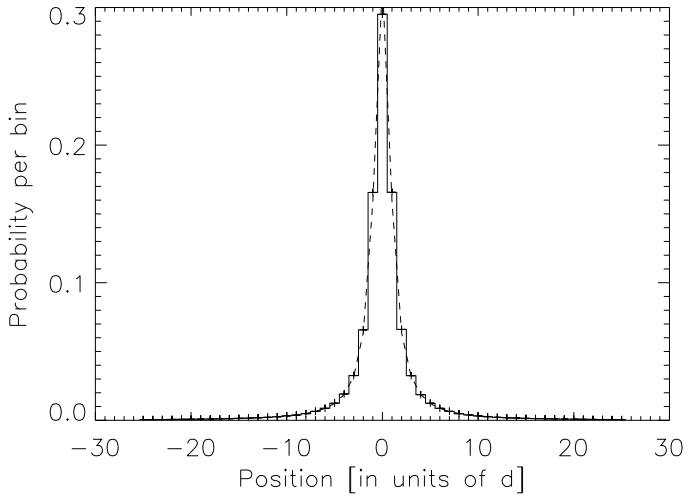

Figure 4.4: The data from Fig. 4.2 transformed from θ to x using Eq. (4.14) and expressed as an observed probabilty. The uniform distribution in θ has been converted into the highly peaked distribution, which is given by Eq. (4.18). The dashed line plots this theoretical distribution. The bins are one unit of d wide.

Chapter 5

SUMMARY

Systematic error Reproducible inaccuracy introduced by faulty equipment, calibration, or technique.
Random error Indefiniteness, of result introduced by finite precision of measurement. Measure of fluctuation after repeated experimentation.
Uncertainty Magnitude of error that is estimated to have been made in determination of results.
Accuracy Measure of how close the result of an experiment comes to the "true" value.
Precision Measure of how carefully the result is determined without reference to any "true" value.
Parent population Hypothetical infinite set of data points of which the experimental data points are assumed to be a random sample.
Parent distribution Probability distribution of the parent population from which the sample data are chosen.
Expectation value Weighted average of a function $f(x)$ over all values of x, $\langle f(x)\rangle = \lim_{N\to\infty} \sum f(x_i)/N = \sum f(x_j)P(x_j) = \int f(x)P(x)dx$
Mean $\mu = \langle x\rangle$
Variance $\sigma^2 = \langle (x_i - \mu)^2\rangle = \langle x^2\rangle - \mu^2$
Standard deviation $\sigma = \sqrt{\sigma^2}$
Sample mean $\bar{x} = \sum x_i/N$
Sample variance $s^2 = \sum (x_i - \bar{x})^2/(N-1)$

Chapter 6

IDL Code

6.1 Accuracy & Precision

Here is the IDL code that shows the difference between accuracy and precision.

```
;-----------------------------------------------
; Illustrate the difference between precision and
; accuracy
;
; JRG 99/8/13
;-----------------------------------------------

!p.multi=[0,2,2] ; set up to make a grid of 2x2 plots

; now simulate the data

nx = 256 ; this will be the number of points
ny = nx

x = randomn(seed,nx) ; generate a normally distributed set of numbers
y = randomn(seed,ny) ; for the x and y coordinates

;-----------------------------------------------
; plot the data
```

```
plot,x,y,ps=3,xr=[-3,3],yr=[-3,3],$
title='Accurate & Imprecise',xthick=2,ythick=2,$
thick=2,xtit='X',ytit='Y',charsize=1.4,charthick=2

; make a bull's eye target

theta = findgen(360)         ; theta is an array 0, 1, 2, ... 359
r = 1.0         ; radius of the circle
xx = r*cos(theta/!radeg) ; use converstion from polar to
yy = r*sin(theta/!radeg) ; cartesian coordinates
oplot,xx,yy,thick=2 ; plot the circle

;-----------------------------------------------
; replot the data with a smaller spread

plot,x*0.5,y*0.5,ps=3,xr=[-3,3],yr=[-3,3],$
title='Accurate & Precise',xthick=2,ythick=2,$
thick=2,xtit='X',ytit='Y',charsize=1.4,charthick=2
oplot,xx,yy,thick=2 ; plot the circle

;-----------------------------------------------
; or with an offset

plot,x+1,y+1,ps=3,xr=[-3,3],yr=[-3,3],$
title='Inaccurate & Imprecise',xthick=2,ythick=2,$
thick=2,xtit='X',ytit='Y',charsize=1.4,charthick=2
oplot,xx,yy,thick=2 ; plot the circle

;-----------------------------------------------
; and finally with a small spread and offset

plot,x*0.5+1,y*0.5+1,ps=3,xr=[-3,3],yr=[-3,3],$
title='Inaccurate & Precise',xthick=2,ythick=2,$
thick=2,xtit='X',ytit='Y',charsize=1.4,charthick=2
oplot,xx,yy,thick=2 ; plot the circle

; all done
```

```
!p.multi=0 ; reset to normal plotting

end
```

6.2 Simulated Experiment

Here is the IDL code that makes the figure illustrating the experiment of measuring lengths.

```
; Simulate an experiment with a Gaussian
; error distribution. Emphasize the difference between
; the parent and the sample populations.
;
; Original: JRG 99/8/13
; Revised:  JRG 06/8/29
;-----------------------------------------------
; The parent mean and standard deviation are

mean = 20.0
sigma = 0.5

; the number of measurements made to be made

nb = 100 ; Try an experiment with smaller and larger samples

; Use the built-in normal distribution generator to
; simulate the experiment

seed = -1 ; change this number to get a different random sequence

h = mean + sigma*randomn(seed,nb)   ; use the built in random function

; Now make a histogram of the experiment by choosing the bin
; size
```

```
binsize = 0.2

min = 18    ; Center of the lowest bin
max = 22    ; Center of highest bin

nbin = (max - min)/binsize + 1   ; number of bins between min and max
hist = intarr(nbin)              ; create an array to save the histogram

; Make the list of bin centers

xhist = findgen(nbin)*binsize + min
;-------------------------------------------
; count up the number of data points in each
; data bin by looping over each bin center

for i=0,nbin-1 do begin

   lo = xhist[i] - binsize/2.
   hi = lo + binsize
   w = where( h ge lo and h lt hi,count)   ; WHERE is one of IDL's most
                                           ; useful functions.
   if count eq -1 then hist[i] = 0 else hist[i] = count

endfor
;-------------------------------------------
; Plot the answer
;
; First plot the bin centers

plot,xhist,hist,ps=1,title = 'Simulated data',xtit='Length (cm)',$
ytit = 'Number of measurements',$
charsize =2,charthick=2,xthick=2,ythick=2,thick=2

; And plot again using the histogram style line

oplot, xhist,hist,ps=10,thick=2
```

```
; Calculate the mean and standard deviation

x = total(h)/nb ; the total function simply adds up all the values in the array
print,'The mean of the sample is',x

; Calculate the standard deviation

s = sqrt(total( (h - mean)^2.0)/(nb - 1.0) )
print,'The standard deviation is',s

; Label the plot

xyouts,0.2,0.85,'Sample x ='+string(x,form='(f6.2)'),charsize=1.4,/normal
xyouts,0.2,0.8,'Sample s ='+string(s,form='(f6.2)'),charsize=1.4,/normal

; now overplot the normal distributions inferred from the sample

xscale = min + findgen(nb)*(max - min)/nb

nd =  nb*binsize*exp (-0.5 * ((xscale - x)/s)^2)/(s*sqrt(2*!pi))
oplot,xscale,nd

; and now the parent (original) distribution

nd =  nb*binsize*exp (-0.5 * ((xscale - mean)/sigma)^2)/(sigma*sqrt(2*!pi))
oplot,xscale,nd,line=1

xyouts,0.7,0.85,'Parent !4l!3 ='+string(mean,form='(f6.2)'),charsize=1.4,/normal
xyouts,0.7,0.8,'Parent !4r!3 ='+string(sigma,form='(f6.2)'),charsize=1.4,/normal

end
```

6.3 Lighthouse

```
; Simulate random flashes from the
; lighthouse using IDL's built in
; random number generator

;nflash = 1024      ; number of flashes
nflash = 2L^20
thetamin = -90.0   ; minimum angle
thetamax = 90.0    ; maximum angle

iseed = -1         ; seed for random number generator

; Make a random number between 0 and 1

nr = randomu(iseed, nflash)     ; randomu is IDL's random number
                                ; funtion

; Convert the random number 0 <= nr <= 1.0

thetar = (nr - 0.5)*(thetamax - thetamin)

; Make a histogram of the angles. First decide on the bin width

binw = 1.0     ; 1 degree wide bins

; The number of bins

nbin = (thetamax - thetamin)/binw + 1.0

bins = thetamin + findgen(nbin)*binw

histocount = fltarr(nbin)

for i=0, nbin-1 do begin
; count up the number of theta in that bin

   w=where(thetar ge bins[i]-binw/2 $
           and thetar lt bins[i]+binw/2,countw)
   histocount[i] = countw
```

```
endfor

plot,bins,histocount/nflash,$
  ps=1,xthick=2,ythick=2,thick=2,charsize=2,charthick=2,$
  xtitle = 'Angle [degrees]',ytitle='Probability per bin',$
  yr=[0,0.016]

oplot,bins,histocount/nflash,ps=10

; Plot a horizantal line at the expected average
; number per bin

nave = 1./nbin

oplot,[-90,90],[nave,nave],line=2,thick=2

;-----------------------------------------

dist = 1.0 ; distance from the shore

xp = dist * tan(thetar/!radeg)

binxw = 1.0     ; 1 degree wide bins

xmin =-25.
xmax = 25.

; The number of bins

nbinx = (xmax - xmin)/binxw + 1.0

xbins = xmin + findgen(nbinx)*binxw

histocount = fltarr(nbinx)

for i=0, nbinx-1 do begin
```

```
; count up the number of theta in that bin

   w = where( xp ge xbins[i]-binxw/2 and xp lt xbins[i]+binxw/2, countw)
   histocount[i] = countw

endfor

plot,xbins,histocount/nflash,$
  ps=1,xthick=2,ythick=2,thick=2,charsize=2,charthick=2,$
  xtitle = 'Position [in units of d]',ytitle='Probability per bin'

oplot,xbins,histocount/nflash,ps=10

oplot, xbins,(1/!pi) /(dist + xbins^2/dist),line=2

end
```