►*Calculate* $\sum_{i=1}^{N} x_i$ *and* $\sum_{i=1}^{N} x_i^2$ *for the data given in table 31.1 and hence find the mean and standard deviation of the sample.*

From table 31.1, we obtain

$$\sum_{i=1}^{N} x_i = 188.7 + 204.7 + \cdots + 200.0 = 1479.8,$$

$$\sum_{i=1}^{N} x_i^2 = (188.7)^2 + (204.7)^2 + \cdots + (200.0)^2 = 275\,334.36.$$

Since $N = 8$, we find as before (quoting the final results to one decimal place)

$$\bar{x} = \frac{1479.8}{8} = 185.0, \qquad s = \sqrt{\frac{275\,334.36}{8} - \left(\frac{1479.8}{8}\right)^2} = 14.2. \blacktriangleleft$$

31.2.3 Moments and central moments

By analogy with our discussion of probability distributions in section 30.5, the sample mean and variance may also be described respectively as the first moment and second central moment of the sample. In general, for a sample x_i, $i = 1, 2, \ldots, N$, we define the rth moment m_r and rth central moment n_r as

$$m_r = \frac{1}{N}\sum_{i=1}^{N} x_i^r, \tag{31.9}$$

$$n_r = \frac{1}{N}\sum_{i=1}^{N} (x_i - m_1)^r. \tag{31.10}$$

Thus the sample mean $\bar{x}$ and variance s^2 may also be written as m_1 and n_2 respectively. As is common practice, we have introduced a notation in which a sample statistic is denoted by the Roman letter corresponding to whichever Greek letter is used to describe the corresponding population statistic. Thus, we use m_r and n_r to denote the rth moment and central moment of a sample, since in section 30.5 we denoted the rth moment and central moment of a population by μ_r and ν_r respectively.

This notation is particularly useful, since the rth central moment of a sample, m_r, may be expressed in terms of the rth- and lower-order sample moments n_r in a way exactly analogous to that derived in subsection 30.5.5 for the corresponding population statistics. As discussed in the previous section, the sample variance is given by $s^2 = \overline{x^2} - \bar{x}^2$ but this may also be written as $n_2 = m_2 - m_1^2$, which is to be compared with the corresponding relation $\nu_2 = \mu_2 - \mu_1^2$ derived in subsection 30.5.3 for population statistics. This correspondence also holds for higher-order central moments of the sample. For example,

$$\begin{aligned} n_3 &= \frac{1}{N}\sum_{i=1}^{N}(x_i - m_1)^3 \\ &= \frac{1}{N}\sum_{i=1}^{N}(x_i^3 - 3m_1x_i^2 + 3m_1^2x_i - m_1^3) \\ &= m_3 - 3m_1m_2 + 3m_1^2m_1 - m_1^3 \\ &= m_3 - 3m_1m_2 + 2m_1^3, \end{aligned} \tag{31.11}$$

which may be compared with equation (30.53) in the previous chapter.

Mirroring our discussion of the normalised central moments γ_r of a population in subsection 30.5.5, we can also describe a sample in terms of the dimensionless quantities

$$g_k = \frac{n_k}{n_2^{k/2}} = \frac{n_k}{s^k};$$

g_3 and g_4 are called the sample skewness and kurtosis. Likewise, it is common to define the *excess* kurtosis of a sample by $g_4 - 3$.

31.2.4 Covariance and correlation

So far we have assumed that each data item of the sample consists of a single number. Now let us suppose that each item of data consists of a pair of numbers, so that the sample is given by (x_i, y_i), $i = 1, 2, \ldots, N$.

We may calculate the sample means, $\bar{x}$ and $\bar{y}$, and sample variances, s_x^2 and s_y^2, of the x_i and y_i values individually but these statistics do not provide any measure of the relationship between the x_i and y_i. By analogy with our discussion in subsection 30.12.3 we measure any interdependence between the x_i and y_i in terms of the *sample covariance*, which is given by

$$\begin{aligned} V_{xy} &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) \\ &= \overline{(x - \bar{x})(y - \bar{y})} \\ &= \overline{xy} - \bar{x}\bar{y}. \end{aligned} \tag{31.12}$$

Writing out the last expression in full, we obtain the form most useful for calculations, which reads

$$V_{xy} = \frac{1}{N}\left(\sum_{i=1}^{N} x_iy_i\right) - \frac{1}{N^2}\left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right).$$

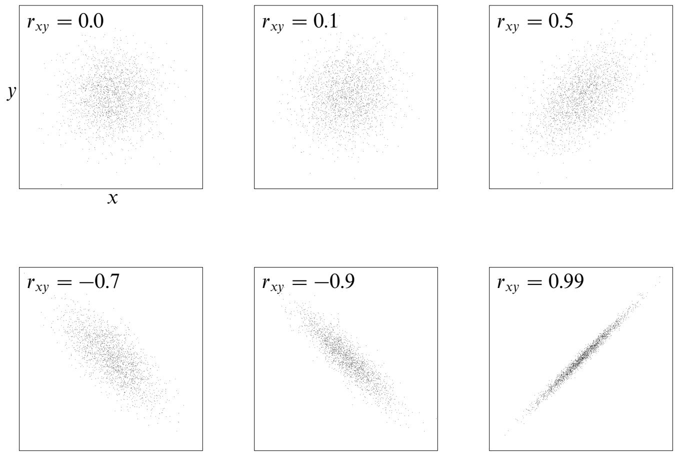


Figure 31.1 Scatter plots for two-dimensional data samples of size $N = 1000$, with various values of the correlation r. No scales are plotted, since the value of r is unaffected by shifts of origin or changes of scale in x and y.

We may also define the closely related *sample correlation* by

$$r_{xy} = \frac{V_{xy}}{s_x s_y},$$

which can take values between -1 and $+1$. If the x_i and y_i are independent then $V_{xy} = 0 = r_{xy}$, and from (31.12) we see that $\overline{xy} = \bar{x}\bar{y}$. It should also be noted that the value of r_{xy} is not altered by shifts in the origin or by changes in the scale of the x_i or y_i. In other words, if $x' = ax + b$ and $y' = cy + d$, where a, b, c, d are constants, then $r_{x'y'} = r_{xy}$. Figure 31.1 shows scatter plots for several two-dimensional random samples x_i, y_i of size $N = 1000$, each with a different value of r_{xy}.

►*Ten UK citizens are selected at random and their heights and weights are found to be as follows (to the nearest* cm *or* kg *respectively):*

Person	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Height (cm)	194	168	177	180	171	190	151	169	175	182
Weight (kg)	75	53	72	80	75	75	57	67	46	68

Calculate the sample correlation between the heights and weights.

In order to find the sample correlation, we begin by calculating the following sums (where x_i are the heights and y_i are the weights)

$$\sum_i x_i = 1757, \qquad \sum_i y_i = 668,$$

$$\sum_i x_i^2 = 310\,041, \qquad \sum_i y_i^2 = 45\,746, \qquad \sum_i x_i y_i = 118\,029.$$

The sample consists of $N = 10$ pairs of numbers, so the means of the x_i and of the y_i are given by $\bar{x} = 175.7$ and $\bar{y} = 66.8$. Also, $\overline{xy} = 11\,802.9$. Similarly, the standard deviations of the x_i and y_i are calculated, using (31.8), as

$$s_x = \sqrt{\frac{310\,041}{10} - \left(\frac{1757}{10}\right)^2} = 11.6,$$

$$s_y = \sqrt{\frac{45\,746}{10} - \left(\frac{668}{10}\right)^2} = 10.6.$$

Thus the sample correlation is given by

$$r_{xy} = \frac{\overline{xy} - \bar{x}\bar{y}}{s_x s_y} = \frac{11\,802.9 - (175.7)(66.8)}{(11.6)(10.6)} = 0.54.$$

Thus there is a moderate positive correlation between the heights and weights of the people measured. ◄

It is straightforward to generalise the above discussion to data samples of arbitrary dimension, the only complication being one of notation. We choose to denote the ith data item from an n-dimensional sample as $(x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(n)})$, where the bracketted superscript runs from 1 to n and labels the elements within a given data item whereas the subscript i runs from 1 to N and labels the data items within the sample. In this n-dimensional case, we can define the *sample covariance matrix* whose elements are

$$V_{kl} = \overline{x^{(k)}x^{(l)}} - \overline{x^{(k)}}\;\overline{x^{(l)}}$$

and the *sample correlation matrix* with elements

$$r_{kl} = \frac{V_{kl}}{s_k s_l}.$$

Both these matrices are clearly symmetric but are *not* necessarily positive definite.

31.3 Estimators and sampling distributions

In general, the population $P(\mathbf{x})$ from which a sample $x_1, x_2, \ldots, x_N$ is drawn is *unknown*. The *central aim* of statistics is to use the sample values x_i to infer certain properties of the unknown population $P(\mathbf{x})$, such as its mean, variance and higher moments. To keep our discussion in general terms, let us denote the various parameters of the population by $a_1, a_2, \ldots$, or collectively by $\mathbf{a}$. Moreover, we make the dependence of the population on the values of these quantities explicit by writing the population as $P(\mathbf{x}|\mathbf{a})$. For the moment, we are assuming that the sample values x_i are independent and drawn from the same (one-dimensional) population $P(x|\mathbf{a})$, in which case

$$P(\mathbf{x}|\mathbf{a}) = P(x_1|\mathbf{a})P(x_2|\mathbf{a}) \cdots P(x_N|\mathbf{a}).$$

Suppose, we wish to *estimate* the value of one of the quantities $a_1, a_2, \ldots,$ which we will denote simply by a. Since the sample values x_i provide our only source of information, any estimate of a must be some function of the x_i, i.e. some sample statistic. Such a statistic is called an *estimator* of a and is usually denoted by $\hat{a}(\mathbf{x})$, where $\mathbf{x}$ denotes the sample elements $x_1, x_2, \ldots, x_N$.

Since an estimator $\hat{a}$ is a function of the sample values of the random variables $x_1, x_2, \ldots, x_N$, it too must be a random variable. In other words, if a number of random samples, each of the same size N, are taken from the (one-dimensional) population $P(x|\mathbf{a})$ then the value of the estimator $\hat{a}$ will vary from one sample to the next and in general will not be equal to the true value a. This variation in the estimator is described by its *sampling distribution* $P(\hat{a}|\mathbf{a})$. From section 30.14, this is given by

$$P(\hat{a}|\mathbf{a})\, d\hat{a} = P(\mathbf{x}|\mathbf{a})\, d^N\mathbf{x},$$

where $d^N\mathbf{x}$ is the infinitesimal 'volume' in $\mathbf{x}$-space lying between the 'surfaces' $\hat{a}(\mathbf{x}) = \hat{a}$ and $\hat{a}(\mathbf{x}) = \hat{a} + d\hat{a}$. The form of the sampling distribution generally depends upon the estimator under consideration and upon the form of the population from which the sample was drawn, including, as indicated, the true values of the quantities $\mathbf{a}$. It is also usually dependent on the sample size N.

►*The sample values $x_1, x_2, \ldots, x_N$ are drawn independently from a Gaussian distribution with mean μ and variance σ. Suppose that we choose the sample mean $\bar{x}$ as our estimator $\hat{\mu}$ of the population mean. Find the sampling distributions of this estimator.*

The sample mean $\bar{x}$ is given by

$$\bar{x} = \frac{1}{N}(x_1 + x_2 + \cdots + x_N),$$

where the x_i are independent random variables distributed as $x_i \sim N(\mu, \sigma^2)$. From our discussion of multiple Gaussian distributions on page 1189, we see immediately that $\bar{x}$ will also be Gaussian distributed as $N(\mu, \sigma^2/N)$. In other words, the sampling distribution of $\bar{x}$ is given by

$$P(\bar{x}|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\frac{(\bar{x}-\mu)^2}{2\sigma^2/N}\right]. \tag{31.13}$$

Note that the variance of this distribution is σ^2/N. ◄

31.3.1 Consistency, bias and efficiency of estimators

For any particular quantity a, we may in fact define any number of different estimators, each of which will have its own sampling distribution. The quality of a given estimator $\hat{a}$ may be assessed by investigating certain properties of its sampling distribution $P(\hat{a}|\mathbf{a})$. In particular, an estimator $\hat{a}$ is usually judged on the three criteria of *consistency*, *bias* and *efficiency*, each of which we now discuss.

Consistency

An estimator $\hat{a}$ is *consistent* if its value tends to the true value a in the large-sample limit, i.e.

$$\lim_{N\to\infty} \hat{a} = a.$$

Consistency is usually a minimum requirement for a useful estimator. An equivalent statement of consistency is that in the limit of large N the sampling distribution $P(\hat{a}|\mathbf{a})$ of the estimator must satisfy

$$\lim_{N\to\infty} P(\hat{a}|\mathbf{a}) \to \delta(\hat{a} - a).$$

Bias

The expectation value of an estimator $\hat{a}$ is given by

$$E[\hat{a}] = \int \hat{a} P(\hat{a}|\mathbf{a})\, d\hat{a} = \int \hat{a}(\mathbf{x}) P(\mathbf{x}|\mathbf{a})\, d^N\mathbf{x}, \qquad (31.14)$$

where the second integral extends over all possible values that can be taken by the sample elements $x_1, x_2, \ldots, x_N$. This expression gives the expected mean value of $\hat{a}$ from an infinite number of samples, each of size N. The *bias* of an estimator $\hat{a}$ is then defined as

$$b(\mathbf{a}) = E[\hat{a}] - a. \qquad (31.15)$$

We note that the bias b does not depend on the measured sample values $x_1, x_2, \ldots, x_N$. In general, though, it will depend on the sample size N, the functional form of the estimator $\hat{a}$ and, as indicated, on the true properties $\mathbf{a}$ of the population, including the true value of a itself. If $b = 0$ then $\hat{a}$ is called an *unbiased* estimator of a.

►*An estimator $\hat{a}$ is biased in such a way that $E[\hat{a}] = a + b(a)$, where the bias $b(a)$ is given by $(b_1 - 1)a + b_2$ and b_1 and b_2 are known constants. Construct an unbiased estimator of a.*

Let us first write $E[\hat{a}]$ is the clearer form

$$E[\hat{a}] = a + (b_1 - 1)a + b_2 = b_1 a + b_2.$$

The task of constructing an unbiased estimator is now trivial, and an appropriate choice is $\hat{a}' = (\hat{a} - b_2)/b_1$, which (as required) has the expectation value

$$E[\hat{a}'] = \frac{E[\hat{a}] - b_2}{b_1} = a. \blacktriangleleft$$

Efficiency

The variance of an estimator is given by

$$V[\hat{a}] = \int (\hat{a} - E[\hat{a}])^2 P(\hat{a}|\mathbf{a})\, d\hat{a} = \int (\hat{a}(\mathbf{x}) - E[\hat{a}])^2 P(\mathbf{x}|\mathbf{a})\, d^N\mathbf{x} \qquad (31.16)$$

and describes the spread of values $\hat{a}$ about $E[\hat{a}]$ that would result from a large number of samples, each of size N. An estimator with a smaller variance is said to be more *efficient* than one with a larger variance. As we show in the next section, for any given quantity a of the population there exists a theoretical *lower limit* on the variance of *any* estimator $\hat{a}$. This result is known as *Fisher's inequality* (or the *Cramér–Rao inequality*) and reads

$$V[\hat{a}] \geq \left(1 + \frac{\partial b}{\partial a}\right)^2 \Big/ E\left[-\frac{\partial^2 \ln P}{\partial a^2}\right], \tag{31.17}$$

where P stands for the population $P(\mathbf{x}|\mathbf{a})$ and b is the bias of the estimator. Denoting the quantity on the RHS of (31.17) by $V_{\min}$, the *efficiency* e of an estimator is defined as

$$e = V_{\min}/V[\hat{a}].$$

An estimator for which $e = 1$ is called a *minimum-variance* or *efficient* estimator. Otherwise, if $e < 1$, $\hat{a}$ is called an *inefficient* estimator.

It should be noted that, in general, there is no unique 'optimal' estimator $\hat{a}$ for a particular property a. To some extent, there is always a trade-off between bias and efficiency. One must often weigh the relative merits of an unbiased, inefficient estimator against another that is more efficient but slightly biased. Nevertheless, a common choice is the *best unbiased estimator* (BUE), which is simply the unbiased estimator $\hat{a}$ having the smallest variance $V[\hat{a}]$.

Finally, we note that some qualities of estimators are related. For example, suppose that $\hat{a}$ is an unbiased estimator, so that $E[\hat{a}] = a$ and $V[\hat{a}] \to 0$ as $N \to \infty$. Using the Bienaymé–Chebyshev inequality discussed in subsection 30.5.3, it follows immediately that $\hat{a}$ is also a consistent estimator. Nevertheless, it does *not* follow that a consistent estimator is unbiased.

►*The sample values $x_1, x_2, \ldots, x_N$ are drawn independently from a Gaussian distribution with mean μ and variance σ. Show that the sample mean $\bar{x}$ is a consistent, unbiased, minimum-variance estimator of μ.*

We found earlier that the sampling distribution of $\bar{x}$ is given by

$$P(\bar{x}|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\frac{(\bar{x}-\mu)^2}{2\sigma^2/N}\right],$$

from which we see immediately that $E[\bar{x}] = \mu$ and $V[\bar{x}] = \sigma^2/N$. Thus $\bar{x}$ is an unbiased estimator of μ. Moreover, since it is also true that $V[\bar{x}] \to 0$ as $N \to \infty$, $\bar{x}$ is a consistent estimator of μ.

In order to determine whether $\bar{x}$ is a minimum-variance estimator of μ, we must use Fisher's inequality (31.17). Since the sample values x_i are independent and drawn from a Gaussian of mean μ and standard deviation σ, we have

$$\ln P(\mathbf{x}|\mu,\sigma) = -\frac{1}{2}\sum_{i=1}^{N}\left[\ln(2\pi\sigma^2) + \frac{(x_i-\mu)^2}{\sigma^2}\right],$$

and, on differentiating twice with respect to μ, we find

$$\frac{\partial^2 \ln P}{\partial \mu^2} = -\frac{N}{\sigma^2}.$$

This is independent of the x_i and so its expectation value is also equal to $-N/\sigma^2$. With b set equal to zero in (31.17), Fisher's inequality thus states that, for *any* unbiased estimator $\hat{\mu}$ of the population mean,

$$V[\hat{\mu}] \geq \frac{\sigma^2}{N}.$$

Since $V[\bar{x}] = \sigma^2/N$, the sample mean $\bar{x}$ is a minimum-variance estimator of μ. ◄

31.3.2 Fisher's inequality

As mentioned above, Fisher's inequality provides a lower limit on the variance of *any* estimator $\hat{a}$ of the quantity a; it reads

$$V[\hat{a}] \geq \left(1 + \frac{\partial b}{\partial a}\right)^2 \Big/ E\left[-\frac{\partial^2 \ln P}{\partial a^2}\right], \qquad (31.18)$$

where P stands for the population $P(\mathbf{x}|\mathbf{a})$ and b is the bias of the estimator. We now present a proof of this inequality. Since the derivation is somewhat complicated, and many of the details are unimportant, this section can be omitted on a first reading. Nevertheless, some aspects of the proof will be useful when the efficiency of maximum-likelihood estimators is discussed in section 31.5.

►*Prove Fisher's inequality (31.18).*

The normalisation of $P(\mathbf{x}|\mathbf{a})$ is given by

$$\int P(\mathbf{x}|\mathbf{a})\, d^N\mathbf{x} = 1, \qquad (31.19)$$

where $d^N\mathbf{x} = dx_1 dx_2 \cdots dx_N$ and the integral extends over all the allowed values of the sample items x_i. Differentiating (31.19) with respect to the parameter a, we obtain

$$\int \frac{\partial P}{\partial a}\, d^N\mathbf{x} = \int \frac{\partial \ln P}{\partial a} P\, d^N\mathbf{x} = 0. \qquad (31.20)$$

We note that the second integral is simply the expectation value of $\partial \ln P/\partial a$, where the average is taken over all possible samples x_i, $i = 1, 2, \ldots, N$. Further, by equating the two expressions for $\partial E[\hat{a}]/\partial a$ obtained by differentiating (31.15) and (31.14) with respect to a we obtain, dropping the functional dependencies, a second relationship,

$$1 + \frac{\partial b}{\partial a} = \int \hat{a} \frac{\partial P}{\partial a}\, d^N\mathbf{x} = \int \hat{a} \frac{\partial \ln P}{\partial a} P\, d^N\mathbf{x}. \qquad (31.21)$$

Now, multiplying (31.20) by $\alpha(a)$, where $\alpha(a)$ is *any* function of a, and subtracting the result from (31.21), we obtain

$$\int [\hat{a} - \alpha(a)] \frac{\partial \ln P}{\partial a} P\, d^N\mathbf{x} = 1 + \frac{\partial b}{\partial a}.$$

At this point we must invoke the Schwarz inequality proved in subsection 8.1.3. The proof is trivially extended to multiple integrals and shows that for two real functions, $g(\mathbf{x})$ and $h(\mathbf{x})$,

$$\left(\int g^2(\mathbf{x})\,d^N\mathbf{x}\right)\left(\int h^2(\mathbf{x})\,d^N\mathbf{x}\right) \geq \left(\int g(\mathbf{x})h(\mathbf{x})\,d^N\mathbf{x}\right)^2. \qquad (31.22)$$

If we now let $g = [\hat{a} - \alpha(a)]\sqrt{P}$ and $h = (\partial \ln P/\partial a)\sqrt{P}$, we find

$$\left\{\int [\hat{a} - \alpha(a)]^2 P\,d^N\mathbf{x}\right\}\left[\int \left(\frac{\partial \ln P}{\partial a}\right)^2 P\,d^N\mathbf{x}\right] \geq \left(1 + \frac{\partial b}{\partial a}\right)^2.$$

On the LHS, the factor in braces represents the expected spread of $\hat{a}$-values around the point $\alpha(a)$. The minimum value that this integral may take occurs when $\alpha(a) = E[\hat{a}]$. Making this substitution, we recognise the integral as the variance $V[\hat{a}]$, and so obtain the result

$$V[\hat{a}] \geq \left(1 + \frac{\partial b}{\partial a}\right)^2 \left[\int \left(\frac{\partial \ln P}{\partial a}\right)^2 P\,d^N\mathbf{x}\right]^{-1}. \qquad (31.23)$$

We note that the factor in brackets is the expectation value of $(\partial \ln P/\partial a)^2$.

Fisher's inequality is, in fact, often quoted in the form (31.23). We may recover the form (31.18) by noting that on differentiating (31.20) with respect to a we obtain

$$\int \left(\frac{\partial^2 \ln P}{\partial a^2} P + \frac{\partial \ln P}{\partial a}\frac{\partial P}{\partial a}\right) d^N\mathbf{x} = 0.$$

Writing $\partial P/\partial a$ as $(\partial \ln P/\partial a)P$ and rearranging we find that

$$\int \left(\frac{\partial \ln P}{\partial a}\right)^2 P\,d^N\mathbf{x} = -\int \frac{\partial^2 \ln P}{\partial a^2} P\,d^N\mathbf{x}.$$

Substituting this result in (31.23) gives

$$V[\hat{a}] \geq -\left(1 + \frac{\partial b}{\partial a}\right)^2 \left[\int \frac{\partial^2 \ln P}{\partial a^2} P\,d^N\mathbf{x}\right]^{-1}.$$

Since the factor in brackets is the expectation value of $\partial^2 \ln P/\partial a^2$, we have recovered result (31.18). ◄

31.3.3 Standard errors on estimators

For a given sample $x_1, x_2, \ldots, x_N$, we may calculate the value of an estimator $\hat{a}(\mathbf{x})$ for the quantity a. It is also necessary, however, to give some measure of the statistical uncertainty in this estimate. One way of characterising this uncertainty is with the standard deviation of the sampling distribution $P(\hat{a}|\mathbf{a})$, which is given simply by

$$\sigma_{\hat{a}} = (V[\hat{a}])^{1/2}. \qquad (31.24)$$

If the estimator $\hat{a}(\mathbf{x})$ were calculated for a large number of samples, each of size N, then the standard deviation of the resulting $\hat{a}$ values would be given by (31.24). Consequently, $\sigma_{\hat{a}}$ is called the *standard error* on our estimate.

In general, however, the standard error $\sigma_{\hat{a}}$ depends on the true values of some or all of the quantities **a** and they may be unknown. When this occurs, one must substitute estimated values of any unknown quantities into the expression for $\sigma_{\hat{a}}$ in order to obtain an estimated standard error $\hat{\sigma}_{\hat{a}}$. One then quotes the result as

$$a = \hat{a} \pm \hat{\sigma}_{\hat{a}}.$$

►*Ten independent sample values x_i, $i = 1, 2, \ldots, 10$, are drawn at random from a Gaussian distribution with standard deviation $\sigma = 1$. The sample values are as follows (to two decimal places):*

2.22	2.56	1.07	0.24	0.18	0.95	0.73	−0.79	2.09	1.81

Estimate the population mean μ, quoting the standard error on your result.

We have shown in the final worked example of subsection 31.3.1 that, in this case, $\bar{x}$ is a consistent, unbiased, minimum-variance estimator of μ and has variance $V[\bar{x}] = \sigma^2/N$. Thus, our estimate of the population mean with its associated standard error is

$$\hat{\mu} = \bar{x} \pm \frac{\sigma}{\sqrt{N}} = 1.11 \pm 0.32.$$

If the true value of σ had not been known, we would have needed to use an estimated value $\hat{\sigma}$ in the expression for the standard error. Useful basic estimators of σ are discussed in subsection 31.4.2. ◄

It should be noted that the above approach is most meaningful for unbiased estimators. In this case, $E[\hat{a}] = a$ and so $\sigma_{\hat{a}}$ describes the spread of $\hat{a}$-values about the true value a. For a biased estimator, however, the spread about the true value a is given by the *root mean square error* $\epsilon_{\hat{a}}$, which is defined by

$$\begin{aligned}
\epsilon_{\hat{a}}^2 &= E[(\hat{a} - a)^2] \\
&= E[(\hat{a} - E[\hat{a}])^2] + (E[\hat{a}] - a)^2 \\
&= V[\hat{a}] + b(\mathbf{a})^2.
\end{aligned}$$

We see that $\epsilon_{\hat{a}}^2$ is the sum of the variance of $\hat{a}$ and the square of the bias and so can be interpreted as the sum of squares of statistical and systematic errors. For a biased estimator, it is often more appropriate to quote the result as

$$a = \hat{a} \pm \epsilon_{\hat{a}}.$$

As above, it may be necessary to use estimated values $\hat{\mathbf{a}}$ in the expression for the root mean square error and thus to quote only an estimate $\hat{\epsilon}_{\hat{a}}$ of the error.

31.3.4 Confidence limits on estimators

An alternative (and often equivalent) way of quoting a statistical error is with a *confidence interval*. Let us assume that, *other* than the quantity of interest a, the quantities **a** have known fixed values. Thus we denote the sampling distribution

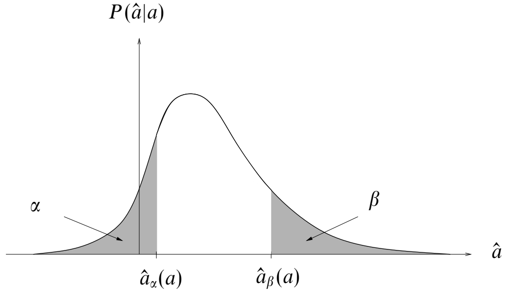


Figure 31.2 The sampling distribution $P(\hat{a}|a)$ of some estimator $\hat{a}$ for a given value of a. The shaded regions indicate the two probabilities $\Pr(\hat{a} < \hat{a}_\alpha(a)) = \alpha$ and $\Pr(\hat{a} > \hat{a}_\beta(a)) = \beta$.

of $\hat{a}$ by $P(\hat{a}|a)$. For any particular value of a, one can determine the two values $\hat{a}_\alpha(a)$ and $\hat{a}_\beta(a)$ such that

$$\Pr(\hat{a} < \hat{a}_\alpha(a)) = \int_{-\infty}^{\hat{a}_\alpha(a)} P(\hat{a}|a)\, d\hat{a} = \alpha, \tag{31.25}$$

$$\Pr(\hat{a} > \hat{a}_\beta(a)) = \int_{\hat{a}_\beta(a)}^{\infty} P(\hat{a}|a)\, d\hat{a} = \beta. \tag{31.26}$$

This is illustrated in figure 31.2. Thus, for any particular value of a, the probability that the estimator $\hat{a}$ lies within the limits $\hat{a}_\alpha(a)$ and $\hat{a}_\beta(a)$ is given by

$$\Pr(\hat{a}_\alpha(a) < \hat{a} < \hat{a}_\beta(a)) = \int_{\hat{a}_\alpha(a)}^{\hat{a}_\beta(a)} P(\hat{a}|a)\, d\hat{a} = 1 - \alpha - \beta.$$

Now, let us suppose that from our sample $x_1, x_2, \ldots, x_N$, we actually obtain the value $\hat{a}_{\rm obs}$ for our estimator. If $\hat{a}$ is a good estimator of a then we would expect $\hat{a}_\alpha(a)$ and $\hat{a}_\beta(a)$ to be monotonically increasing functions of a (i.e. $\hat{a}_\alpha$ and $\hat{a}_\beta$ *both* change in the *same* sense as a when the latter is varied). Assuming this to be the case, we can uniquely define the two numbers a_- and a_+ by the relationships

$$\hat{a}_\alpha(a_+) = \hat{a}_{\rm obs} \qquad \text{and} \qquad \hat{a}_\beta(a_-) = \hat{a}_{\rm obs}.$$

From (31.25) and (31.26) it follows that

$$\Pr(a_+ < a) = \alpha \qquad \text{and} \qquad \Pr(a_- > a) = \beta,$$

which when taken together imply

$$\Pr(a_- < a < a_+) = 1 - \alpha - \beta. \tag{31.27}$$

Thus, from our estimate $\hat{a}_{\rm obs}$, we have determined two values a_- and a_+ such that this interval contains the true value of a with probability $1 - \alpha - \beta$. It should be emphasised that a_- and a_+ are random variables. If a large number of samples,

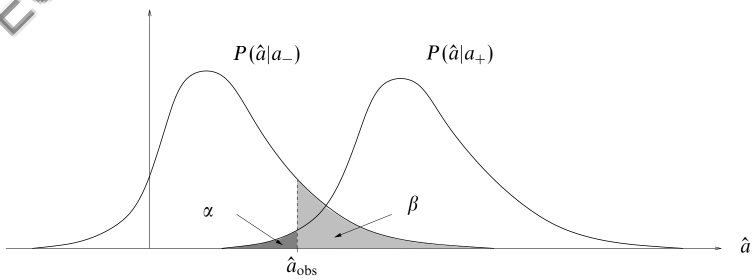


Figure 31.3 An illustration of how the observed value of the estimator, $\hat{a}_{\rm obs}$, and the given values α and β determine the two confidence limits a_- and a_+, which are such that $\hat{a}_\alpha(a_+) = \hat{a}_{\rm obs} = \hat{a}_\beta(a_-)$.

each of size N, were analysed then the interval $[a_-, a_+]$ would contain the true value a on a fraction $1 - \alpha - \beta$ of the occasions.

The interval $[a_-, a_+]$ is called a *confidence interval* on a at the *confidence level* $1 - \alpha - \beta$. The values a_- and a_+ themselves are called respectively the *lower confidence limit* and the *upper confidence limit* at this confidence level. In practice, the confidence level is often quoted as a percentage. A convenient way of presenting our results is

$$\int_{-\infty}^{\hat{a}_{\rm obs}} P(\hat{a}|a_+)\, d\hat{a} = \alpha, \qquad (31.28)$$

$$\int_{\hat{a}_{\rm obs}}^{\infty} P(\hat{a}|a_-)\, d\hat{a} = \beta. \qquad (31.29)$$

The confidence limits may then be found by solving these equations for a_- and a_+ either analytically or numerically. The situation is illustrated graphically in figure 31.3.

Occasionally one might not combine the results (31.28) and (31.29) but use either one or the other to provide a *one-sided* confidence interval on a. Whenever the results are combined to provide a *two-sided* confidence interval, however, the interval is *not* specified uniquely by the confidence level $1-\alpha-\beta$. In other words, there are generally an infinite number of intervals $[a_-, a_+]$ for which (31.27) holds. To specify a unique interval, one often chooses $\alpha = \beta$, resulting in the *central confidence interval* on a. All cases can be covered by calculating the quantities $c = \hat{a} - a_-$ and $d = a_+ - \hat{a}$ and quoting the result of an estimate as

$$a = \hat{a}^{+d}_{-c}.$$

So far we have assumed that the quantities **a** other than the quantity of interest a are known in advance. If this is not the case then the construction of confidence limits is considerably more complicated. This is discussed in subsection 31.3.6.

31.3.5 Confidence limits for a Gaussian sampling distribution

An important special case occurs when the sampling distribution is Gaussian; if the mean is a and the standard deviation is $\sigma_{\hat{a}}$ then

$$P(\hat{a}|a,\sigma_{\hat{a}}) = \frac{1}{\sqrt{2\pi\sigma_{\hat{a}}^2}} \exp\left[-\frac{(\hat{a}-a)^2}{2\sigma_{\hat{a}}^2}\right]. \tag{31.30}$$

For almost any (consistent) estimator $\hat{a}$, the sampling distribution will tend to this form in the large-sample limit $N \to \infty$, as a consequence of the central limit theorem. For a sampling distribution of the form (31.30), the above procedure for determining confidence intervals becomes straightforward. Suppose, from our sample, we obtain the value $\hat{a}_{\text{obs}}$ for our estimator. In this case, equations (31.28) and (31.29) become

$$\Phi\left(\frac{\hat{a}_{\text{obs}} - a_+}{\sigma_{\hat{a}}}\right) = \alpha,$$
$$1 - \Phi\left(\frac{\hat{a}_{\text{obs}} - a_-}{\sigma_{\hat{a}}}\right) = \beta,$$

where $\Phi(z)$ is the cumulative probability function for the standard Gaussian distribution, discussed in subsection 30.9.1. Solving these equations for a_- and a_+ gives

$$a_- = \hat{a}_{\text{obs}} - \sigma_{\hat{a}}\Phi^{-1}(1-\beta), \tag{31.31}$$
$$a_+ = \hat{a}_{\text{obs}} + \sigma_{\hat{a}}\Phi^{-1}(1-\alpha); \tag{31.32}$$

we have used the fact that $\Phi^{-1}(\alpha) = -\Phi^{-1}(1-\alpha)$ to make the equations symmetric. The value of the inverse function $\Phi^{-1}(z)$ can be read off directly from table 30.3, given in subsection 30.9.1. For the normally used central confidence interval one has $\alpha = \beta$. In this case, we see that quoting a result using the standard error, as

$$a = \hat{a} \pm \sigma_{\hat{a}}, \tag{31.33}$$

is equivalent to taking $\Phi^{-1}(1-\alpha) = 1$. From table 30.3, we find $\alpha = 1 - 0.8413 = 0.1587$, and so this corresponds to a confidence level of $1 - 2(0.1587) \approx 0.683$. Thus, the standard error limits give the 68.3% central confidence interval.

►*Ten independent sample values x_i, $i = 1, 2, \ldots, 10$, are drawn at random from a Gaussian distribution with standard deviation $\sigma = 1$. The sample values are as follows (to two decimal places):*

2.22	2.56	1.07	0.24	0.18	0.95	0.73	−0.79	2.09	1.81

Find the 90% central confidence interval on the population mean μ.

Our estimator $\hat{\mu}$ is the sample mean $\bar{x}$. As shown towards the end of section 31.3, the sampling distribution of $\bar{x}$ is Gaussian with mean $E[\bar{x}]$ and variance $V[\bar{x}] = \sigma^2/N$. Since $\sigma = 1$ in this case, the standard error is given by $\sigma_{\hat{x}} = \sigma/\sqrt{N} = 0.32$. Moreover, in subsection 31.3.3, we found the mean of the above sample to be $\bar{x} = 1.11$.

For the 90% central confidence interval, we require $\alpha = \beta = 0.05$. From table 30.3, we find

$$\Phi^{-1}(1-\alpha) = \Phi^{-1}(0.95) = 1.65,$$

and using (31.31) and (31.32) we obtain

$$a_- = \bar{x} - 1.65\sigma_{\bar{x}} = 1.11 - (1.65)(0.32) = 0.58,$$
$$a_+ = \bar{x} + 1.65\sigma_{\bar{x}} = 1.11 + (1.65)(0.32) = 1.64.$$

Thus, the 90% central confidence interval on μ is $[0.58, 1.64]$. For comparison, the true value used to create the sample was $\mu = 1$. ◀

In the case where the standard error $\sigma_{\hat{a}}$ in (31.33) is not known in advance, one must use a value $\hat{\sigma}_{\hat{a}}$ estimated from the sample. In principle, this complicates somewhat the construction of confidence intervals, since properly one should consider the two-dimensional joint sampling distribution $P(\hat{a}, \hat{\sigma}_{\hat{a}}|a)$. Nevertheless, in practice, provided $\hat{\sigma}_{\hat{a}}$ is a fairly good estimate of $\sigma_{\hat{a}}$ the above procedure may be applied with reasonable accuracy. In the special case where the sample values x_i are drawn from a Gaussian distribution with unknown μ and σ, it is in fact possible to obtain *exact* confidence intervals on the mean μ, for a sample of any size N, using Student's t-distribution. This is discussed in subsection 31.7.5.

31.3.6 Estimation of several quantities simultaneously

Suppose one uses a sample $x_1, x_2, \ldots, x_N$ to calculate the values of several estimators $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_M$ (collectively denoted by $\hat{\mathbf{a}}$) of the quantities $a_1, a_2, \ldots, a_M$ (collectively denoted by $\mathbf{a}$) that describe the population from which the sample was drawn. The joint sampling distribution of these estimators is an M-dimensional PDF $P(\hat{\mathbf{a}}|\mathbf{a})$ given by

$$P(\hat{\mathbf{a}}|\mathbf{a})\, d^M\hat{\mathbf{a}} = P(\mathbf{x}|\mathbf{a})\, d^N\mathbf{x}.$$

►*Sample values $x_1, x_2, \ldots, x_N$ are drawn independently from a Gaussian distribution with mean μ and standard deviation σ. Suppose we choose the sample mean $\bar{x}$ and sample standard deviation s respectively as estimators $\hat{\mu}$ and $\hat{\sigma}$. Find the joint sampling distribution of these estimators.*

Since each data value x_i in the sample is assumed to be independent of the others, the joint probability distribution of sample values is given by

$$P(\mathbf{x}|\mu, \sigma) = (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{\sum_i (x_i - \mu)^2}{2\sigma^2}\right].$$

We may rewrite the sum in the exponent as follows:

$$\begin{aligned}\sum_i (x_i - \mu)^2 &= \sum_i (x_i - \bar{x} + \bar{x} - \mu)^2 \\ &= \sum_i (x_i - \bar{x})^2 + 2(\bar{x} - \mu)\sum_i (x_i - \bar{x}) + \sum_i (\bar{x} - \mu)^2 \\ &= Ns^2 + N(\bar{x} - \mu)^2,\end{aligned}$$

where in the last line we have used the fact that $\sum_i (x_i - \bar{x}) = 0$. Hence, for given values of μ and σ, the sampling distribution is in fact a function only of the sample mean $\bar{x}$ and the standard deviation s. Thus the sampling distribution of $\bar{x}$ and s must satisfy

$$P(\bar{x}, s|\mu, \sigma)\, d\bar{x}\, ds = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{N[(\bar{x}-\mu)^2 + s^2]}{2\sigma^2}\right\} dV, \tag{31.34}$$

where $dV = dx_1\, dx_2 \cdots dx_N$ is an element of volume in the sample space which yields simultaneously values of $\bar{x}$ and s that lie within the region bounded by $[\bar{x}, \bar{x} + d\bar{x}]$ and $[s, s + ds]$. Thus our only remaining task is to express dV in terms of $\bar{x}$ and s and their differentials.

Let S be the point in sample space representing the sample $(x_1, x_2, \ldots, x_N)$. For given values of $\bar{x}$ and s, we require the sample values to satisfy both the condition

$$\sum_i x_i = N\bar{x},$$

which defines an $(N-1)$-dimensional hyperplane in the sample space, and the condition

$$\sum_i (x_i - \bar{x})^2 = Ns^2,$$

which defines an $(N-1)$-dimensional hypersphere. Thus S is constrained to lie in the intersection of these two hypersurfaces, which is itself an $(N-2)$-dimensional hypersphere. Now, the volume of an $(N-2)$-dimensional hypersphere is proportional to s^{N-1}. It follows that the volume dV between two concentric $(N-2)$-dimensional hyperspheres of radius $\sqrt{N}s$ and $\sqrt{N}(s + ds)$ and two $(N-1)$-dimensional hyperplanes corresponding to $\bar{x}$ and $\bar{x} + d\bar{x}$ is

$$dV = As^{N-2}\, ds\, d\bar{x},$$

where A is some constant. Thus, substituting this expression for dV into (31.34), we find

$$P(\bar{x}, s|\mu, \sigma) = C_1 \exp\left[-\frac{N(\bar{x}-\mu)^2}{2\sigma^2}\right] C_2 s^{N-2} \exp\left(-\frac{Ns^2}{2\sigma^2}\right) = P(\bar{x}|\mu, \sigma) P(s|\sigma), \tag{31.35}$$

where C_1 and C_2 are constants. We have written $P(\bar{x}, s|\mu, \sigma)$ in this form to show that it separates naturally into two parts, one depending only on $\bar{x}$ and the other only on s. Thus, $\bar{x}$ and s are *independent* variables. Separate normalisations of the two factors in (31.35) require

$$C_1 = \left(\frac{N}{2\pi\sigma^2}\right)^{1/2} \quad \text{and} \quad C_2 = 2\left(\frac{N}{2\sigma^2}\right)^{(N-1)/2} \frac{1}{\Gamma\left(\frac{1}{2}(N-1)\right)},$$

where the calculation of C_2 requires the use of the gamma function, discussed in the Appendix. ◀

The *marginal* sampling distribution of any one of the estimators $\hat{a}_i$ is given simply by

$$P(\hat{a}_i|\mathbf{a}) = \int \cdots \int P(\hat{\mathbf{a}}|\mathbf{a})\, d\hat{a}_1 \cdots d\hat{a}_{i-1} d\hat{a}_{i+1} \cdots d\hat{a}_M,$$

and the expectation value $E[\hat{a}_i]$ and variance $V[\hat{a}_i]$ of $\hat{a}_i$ are again given by (31.14) and (31.16) respectively. By analogy with the one-dimensional case, the standard error $\sigma_{\hat{a}_i}$ on the estimator $\hat{a}_i$ is given by the positive square root of $V[\hat{a}_i]$. With several estimators, however, it is usual to quote their full covariance matrix. This $M \times M$ matrix has elements

$$V_{ij} = \mathrm{Cov}[\hat{a}_i, \hat{a}_j] = \int (\hat{a}_i - E[\hat{a}_i])(\hat{a}_j - E[\hat{a}_j])P(\hat{\mathbf{a}}|\mathbf{a})\, d^M\hat{\mathbf{a}}$$
$$= \int (\hat{a}_i - E[\hat{a}_i])(\hat{a}_j - E[\hat{a}_j])P(\mathbf{x}|\mathbf{a})\, d^N\mathbf{x}.$$

Fisher's inequality can be generalised to the multi-dimensional case. Adapting the proof given in subsection 31.3.2, one may show that, in the case where the estimators are efficient and have zero bias, the elements of the *inverse* of the covariance matrix are given by

$$(V^{-1})_{ij} = E\left[-\frac{\partial^2 \ln P}{\partial a_i \partial a_j}\right], \tag{31.36}$$

where P denotes the population $P(\mathbf{x}|\mathbf{a})$ from which the sample is drawn. The quantity on the RHS of (31.36) is the element F_{ij} of the so-called *Fisher matrix* F of the estimators.

►*Calculate the covariance matrix of the estimators $\bar{x}$ and s in the previous example.*

As shown in (31.35), the joint sampling distribution $P(\bar{x}, s|\mu, \sigma)$ factorises, and so the estimators $\bar{x}$ and s are independent. Thus, we conclude immediately that

$$\mathrm{Cov}[\bar{x}, s] = 0.$$

Since we have already shown in the worked example at the end of subsection 31.3.1 that $V[\bar{x}] = \sigma^2/N$, it only remains to calculate $V[s]$. From (31.35), we find

$$E[s^r] = C_2 \int_0^\infty s^{N-2+r} \exp\left(-\frac{Ns^2}{2\sigma^2}\right) ds = \left(\frac{2}{N}\right)^{r/2} \frac{\Gamma\left(\frac{1}{2}(N-1+r)\right)}{\Gamma\left(\frac{1}{2}(N-1)\right)}\sigma^r,$$

where we have evaluated the integral using the definition of the gamma function given in the Appendix. Thus, the expectation value of the sample standard deviation is

$$E[s] = \left(\frac{2}{N}\right)^{1/2} \frac{\Gamma\left(\frac{1}{2}N\right)}{\Gamma\left(\frac{1}{2}(N-1)\right)}\sigma, \tag{31.37}$$

and its variance is given by

$$V[s] = E[s^2] - (E[s])^2 = \frac{\sigma^2}{N}\left\{N - 1 - 2\left[\frac{\Gamma\left(\frac{1}{2}N\right)}{\Gamma\left(\frac{1}{2}(N-1)\right)}\right]^2\right\}$$

We note, in passing, that (31.37) shows that s is a *biased* estimator of σ. ◄

The idea of a confidence interval can also be extended to the case where several quantities are estimated simultaneously but then the practical construction of an interval is considerably more complicated. The general approach is to construct an M-dimensional *confidence region* R in $\mathbf{a}$-space. By analogy with the one-dimensional case, for a given confidence level of (say) $1 - \alpha$, one first constructs a region $\hat{R}$ in $\hat{\mathbf{a}}$-space, such that

$$\int\!\!\int_{\hat{R}} P(\hat{\mathbf{a}}|\mathbf{a})\, d^M\hat{\mathbf{a}} = 1 - \alpha.$$

A common choice for such a region is that bounded by the 'surface' $P(\hat{\mathbf{a}}|\mathbf{a}) =$ constant. By considering all possible values $\mathbf{a}$ and the values of $\hat{\mathbf{a}}$ lying within the region $\hat{R}$, one can construct a $2M$-dimensional region in the combined space $(\hat{\mathbf{a}}, \mathbf{a})$. Suppose now that, from our sample $\mathbf{x}$, the values of the estimators are $\hat{a}_{i,\text{obs}}$, $i = 1, 2, \ldots, M$. The intersection of the M 'hyperplanes' $\hat{a}_i = \hat{a}_{i,\text{obs}}$ with the $2M$-dimensional region will determine an M-dimensional region which, when projected onto $\mathbf{a}$-space, will determine a confidence limit R at the confidence level $1 - \alpha$. It is usually the case that this confidence region has to be evaluated numerically.

The above procedure is clearly rather complicated in general and a simpler approximate method that uses the likelihood function is discussed in subsection 31.5.5. As a consequence of the central limit theorem, however, in the large-sample limit, $N \to \infty$, the joint sampling distribution $P(\hat{\mathbf{a}}|\mathbf{a})$ will tend, in general, towards the multivariate Gaussian

$$P(\hat{\mathbf{a}}|\mathbf{a}) = \frac{1}{(2\pi)^{M/2}|\mathsf{V}|^{1/2}} \exp\left[-\tfrac{1}{2}Q(\hat{\mathbf{a}}, \mathbf{a})\right], \tag{31.38}$$

where V is the covariance matrix of the estimators and the quadratic form Q is given by

$$Q(\hat{\mathbf{a}}, \mathbf{a}) = (\hat{\mathbf{a}} - \mathbf{a})^{\mathrm{T}}\mathsf{V}^{-1}(\hat{\mathbf{a}} - \mathbf{a}).$$

Moreover, in the limit of large N, the inverse covariance matrix tends to the Fisher matrix F given in (31.36), i.e. $\mathsf{V}^{-1} \to \mathsf{F}$.

For the Gaussian sampling distribution (31.38), the process of obtaining confidence intervals is greatly simplified. The surfaces of constant $P(\hat{\mathbf{a}}|\mathbf{a})$ correspond to surfaces of constant $Q(\hat{\mathbf{a}}, \mathbf{a})$, which have the shape of M-dimensional ellipsoids in $\hat{\mathbf{a}}$-space, centred on the true values $\mathbf{a}$. In particular, let us suppose that the ellipsoid $Q(\hat{\mathbf{a}}, \mathbf{a}) = c$ (where c is some constant) contains a fraction $1 - \alpha$ of the total probability. Now suppose that, from our sample $\mathbf{x}$, we obtain the values $\hat{\mathbf{a}}_{\text{obs}}$ for our estimators. Because of the obvious symmetry of the quadratic form Q with respect to $\mathbf{a}$ and $\hat{\mathbf{a}}$, it is clear that the ellipsoid $Q(\mathbf{a}, \hat{\mathbf{a}}_{\text{obs}}) = c$ in $\mathbf{a}$-space that is centred on $\hat{\mathbf{a}}_{\text{obs}}$ should contain the true values $\mathbf{a}$ with probability $1 - \alpha$. Thus $Q(\mathbf{a}, \hat{\mathbf{a}}_{\text{obs}}) = c$ defines our required confidence region R at this confidence level. This is illustrated in figure 31.4 for the two-dimensional case.

It remains only to determine the constant c corresponding to the confidence level $1 - \alpha$. As discussed in subsection 30.15.2, the quantity $Q(\hat{\mathbf{a}}, \mathbf{a})$ is distributed as a χ^2 variable of order M. Thus, the confidence region corresponding to the

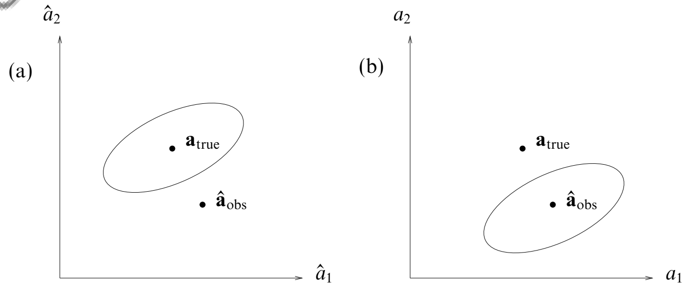


Figure 31.4 (a) The ellipse $Q(\hat{\mathbf{a}}, \mathbf{a}) = c$ in $\hat{\mathbf{a}}$-space. (b) The ellipse $Q(\mathbf{a}, \hat{\mathbf{a}}_{\text{obs}}) = c$ in $\mathbf{a}$-space that corresponds to a confidence region R at the level $1 - \alpha$, when c satisfies (31.39).

confidence level $1 - \alpha$ is given by $Q(\mathbf{a}, \hat{\mathbf{a}}_{\text{obs}}) = c$, where the constant c satisfies

$$\int_0^c P(\chi_M^2)\, d(\chi_M^2) = 1 - \alpha, \tag{31.39}$$

and $P(\chi_M^2)$ is the chi-squared PDF of order M, discussed in subsection 30.9.4. This integral may be evaluated numerically to determine the constant c. Alternatively, some reference books tabulate the values of c corresponding to given confidence levels and various values of M. A representative selection of values of c is given in table 31.2; there the number of degrees of freedom is denoted by the more usual n, rather than M.

31.4 Some basic estimators

In many cases, one does not know the functional form of the population from which a sample is drawn. Nevertheless, in a case where the sample values $x_1, x_2, \ldots, x_N$ are each drawn *independently* from a one-dimensional population $P(x)$, it is possible to construct some basic estimators for the moments and central moments of $P(x)$. In this section, we investigate the estimating properties of the common sample statistics presented in section 31.2. In fact, expectation values and variances of these sample statistics can be calculated *without* prior knowledge of the functional form of the population; they depend only on the sample size N and certain moments and central moments of $P(x)$.

31.4.1 Population mean μ

Let us suppose that the parent population $P(x)$ has mean μ and variance σ^2. An obvious estimator $\hat{\mu}$ of the population mean is the sample mean $\bar{x}$. Provided μ and σ^2 are both finite, we may apply the central limit theorem directly to obtain

%	99	95	10	5	2.5	1	0.5	0.1
$n = 1$	$1.57\,10^{-4}$	$3.93\,10^{-3}$	2.71	3.84	5.02	6.63	7.88	10.83
2	$2.01\,10^{-2}$	0.103	4.61	5.99	7.38	9.21	10.60	13.81
3	0.115	0.352	6.25	7.81	9.35	11.34	12.84	16.27
4	0.297	0.711	7.78	9.49	11.14	13.28	14.86	18.47
5	0.554	1.15	9.24	11.07	12.83	15.09	16.75	20.52
6	0.872	1.64	10.64	12.59	14.45	16.81	18.55	22.46
7	1.24	2.17	12.02	14.07	16.01	18.48	20.28	24.32
8	1.65	2.73	13.36	15.51	17.53	20.09	21.95	26.12
9	2.09	3.33	14.68	16.92	19.02	21.67	23.59	27.88
10	2.56	3.94	15.99	18.31	20.48	23.21	25.19	29.59
11	3.05	4.57	17.28	19.68	21.92	24.73	26.76	31.26
12	3.57	5.23	18.55	21.03	23.34	26.22	28.30	32.91
13	4.11	5.89	19.81	22.36	24.74	27.69	29.82	34.53
14	4.66	6.57	21.06	23.68	26.12	29.14	31.32	36.12
15	5.23	7.26	22.31	25.00	27.49	30.58	32.80	37.70
16	5.81	7.96	23.54	26.30	28.85	32.00	34.27	39.25
17	6.41	8.67	24.77	27.59	30.19	33.41	35.72	40.79
18	7.01	9.39	25.99	28.87	31.53	34.81	37.16	42.31
19	7.63	10.12	27.20	30.14	32.85	36.19	38.58	43.82
20	8.26	10.85	28.41	31.41	34.17	37.57	40.00	45.31
21	8.90	11.59	29.62	32.67	35.48	38.93	41.40	46.80
22	9.54	12.34	30.81	33.92	36.78	40.29	42.80	48.27
23	10.20	13.09	32.01	35.17	38.08	41.64	44.18	49.73
24	10.86	13.85	33.20	36.42	39.36	42.98	45.56	51.18
25	11.52	14.61	34.38	37.65	40.65	44.31	46.93	52.62
30	14.95	18.49	40.26	43.77	46.98	50.89	53.67	59.70
40	22.16	26.51	51.81	55.76	59.34	63.69	66.77	73.40
50	29.71	34.76	63.17	67.50	71.42	76.15	79.49	86.66
60	37.48	43.19	74.40	79.08	83.30	88.38	91.95	99.61
70	45.44	51.74	85.53	90.53	95.02	100.4	104.2	112.3
80	53.54	60.39	96.58	101.9	106.6	112.3	116.3	124.8
90	61.75	69.13	107.6	113.1	118.1	124.1	128.3	137.2
100	70.06	77.93	118.5	124.3	129.6	135.8	140.2	149.4

Table 31.2 The tabulated values are those which a variable distributed as χ^2 with n degrees of freedom exceeds with the given percentage probability. For example, a variable having a χ^2 distribution with 14 degrees of freedom takes values in excess of 21.06 on 10% of occasions.

exact expressions, valid for samples of any size N, for the expectation value and variance of $\bar{x}$. From parts (i) and (ii) of the central limit theorem, discussed in section 30.10, we immediately obtain

$$E[\bar{x}] = \mu, \qquad V[\bar{x}] = \frac{\sigma^2}{N}. \tag{31.40}$$

Thus we see that $\bar{x}$ is an unbiased estimator of μ. Moreover, we note that the standard error in $\bar{x}$ is $\sigma/\sqrt{N}$, and so the sampling distribution of $\bar{x}$ becomes more tightly centred around μ as the sample size N increases. Indeed, since $V[\bar{x}] \to 0$ as $N \to \infty$, $\bar{x}$ is also a consistent estimator of μ.

In the limit of large N, we may in fact obtain an *approximate* form for the full sampling distribution of $\bar{x}$. Part (iii) of the central limit theorem (see section 30.10) tells us immediately that, for large N, the sampling distribution of $\bar{x}$ is given approximately by the Gaussian form

$$P(\bar{x}|\mu,\sigma) \approx \frac{1}{\sqrt{2\pi\sigma^2/N}} \exp\left[-\frac{(\bar{x}-\mu)^2}{2\sigma^2/N}\right].$$

Note that this does *not* depend on the form of the original parent population. If, however, the parent population is in fact Gaussian then this result is *exact* for samples of *any* size N (as is immediately apparent from our discussion of multiple Gaussian distributions in subsection 30.9.1).

31.4.2 Population variance σ^2

An estimator for the population variance σ^2 is not so straightforward to define as one for the mean. Complications arise because, in many cases, the true mean of the population μ is not known. Nevertheless, let us begin by considering the case where in fact μ is known. In this event, a useful estimator is

$$\widehat{\sigma^2} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 = \left(\frac{1}{N}\sum_{i=1}^{N}x_i^2\right) - \mu^2. \tag{31.41}$$

►*Show that $\widehat{\sigma^2}$ is an unbiased and consistent estimator of the population variance σ^2.*

The expectation value of $\widehat{\sigma^2}$ is given by

$$E[\widehat{\sigma^2}] = \frac{1}{N}E\left[\sum_{i=1}^{N}x_i^2\right] - \mu^2 = E[x_i^2] - \mu^2 = \mu_2 - \mu^2 = \sigma^2,$$

from which we see that the estimator is unbiased. The variance of the estimator is

$$V[\widehat{\sigma^2}] = \frac{1}{N^2}V\left[\sum_{i=1}^{N}x_i^2\right] + V[\mu^2] = \frac{1}{N}V[x_i^2] = \frac{1}{N}(\mu_4 - \mu_2^2),$$

in which we have used that fact that $V[\mu^2] = 0$ and $V[x_i^2] = E[x_i^4] - (E[x_i^2])^2 = \mu_4 - \mu_2^2$, where μ_r is the rth population moment. Since $\widehat{\sigma^2}$ is unbiased and $V[\widehat{\sigma^2}] \to 0$ as $N \to \infty$, showing that it is also a consistent estimator of σ^2, the result is established. ◄

If the true mean of the population is unknown, however, a natural alternative is to replace μ by $\bar{x}$ in (31.41), so that our estimator is simply the sample variance s^2 given by

$$s^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2.$$

In order to determine the properties of this estimator, we must calculate $E[s^2]$ and $V[s^2]$. This task is straightforward but lengthy. However, for the investigation of the properties of a *central* moment of the sample, there exists a useful trick that simplifies the calculation. We can assume, with no loss of generality, that the mean μ_1 of the population from which the sample is drawn is equal to zero. With this assumption, the population central moments, ν_r, are identical to the corresponding moments μ_r, and we may perform our calculation in terms of the latter. At the end, however, we replace μ_r by ν_r in the final result and so obtain a general expression that is valid even in cases where $\mu_1 \neq 0$.

►*Calculate $E[s^2]$ and $V[s^2]$ for a sample of size N.*

The expectation value of the sample variance s^2 for a sample of size N is given by

$$E[s^2] = \frac{1}{N}E\left[\sum_i x_i^2\right] - \frac{1}{N^2}E\left[\left(\sum_i x_i\right)^2\right]$$
$$= \frac{1}{N}NE[x_i^2] - \frac{1}{N^2}E\left[\sum_i x_i^2 + \sum_{\substack{i,j\\ j\neq i}} x_i x_j\right]. \qquad (31.42)$$

The number of terms in the double summation in (31.42) is $N(N-1)$, so we find

$$E[s^2] = E[x_i^2] - \frac{1}{N^2}(NE[x_i^2] + N(N-1)E[x_i x_j]).$$

Now, since the sample elements x_i and x_j are independent, $E[x_i x_j] = E[x_i]E[x_j] = 0$, assuming the mean μ_1 of the parent population to be zero. Denoting the rth moment of the population by μ_r, we thus obtain

$$E[s^2] = \mu_2 - \frac{\mu_2}{N} = \frac{N-1}{N}\mu_2 = \frac{N-1}{N}\sigma^2, \qquad (31.43)$$

where in the last line we have used the fact that the population mean is zero, and so $\mu_2 = \nu_2 = \sigma^2$. However, the final result is also valid in the case where $\mu_1 \neq 0$.

Using the above method, we can also find the variance of s^2, although the algebra is rather heavy going. The variance of s^2 is given by

$$V[s^2] = E[s^4] - (E[s^2])^2, \qquad (31.44)$$

where $E[s^2]$ is given by (31.43). We therefore need only consider how to calculate $E[s^4]$, where s^4 is given by

$$s^4 = \left[\frac{\sum_i x_i^2}{N} - \left(\frac{\sum_i x_i}{N} \right)^2 \right]^2$$
$$= \frac{(\sum_i x_i^2)^2}{N^2} - 2\frac{(\sum_i x_i^2)(\sum_i x_i)^2}{N^3} + \frac{(\sum_i x_i)^4}{N^4}. \quad (31.45)$$

We will consider in turn each of the three terms on the RHS. In the first term, the sum $(\sum_i x_i^2)^2$ can be written as

$$\left(\sum_i x_i^2 \right)^2 = \sum_i x_i^4 + \sum_{\substack{i,j \\ j \neq i}} x_i^2 x_j^2,$$

where the first sum contains N terms and the second contains $N(N-1)$ terms. Since the sample elements x_i and x_j are assumed independent, we have $E[x_i^2 x_j^2] = E[x_i^2]E[x_j^2] = \mu_2^2$, and so

$$E\left[\left(\sum_i x_i^2 \right)^2 \right] = N\mu_4 + N(N-1)\mu_2^2.$$

Turning to the second term on the RHS of (31.45),

$$\left(\sum_i x_i^2 \right) \left(\sum_i x_i \right)^2 = \sum_i x_i^4 + \sum_{\substack{i,j \\ j \neq i}} x_i^3 x_j + \sum_{\substack{i,j \\ j \neq i}} x_i^2 x_j^2 + \sum_{\substack{i,j,k \\ k \neq j \neq i}} x_i^2 x_j x_k.$$

Since the mean of the population has been assumed to equal zero, the expectation values of the second and fourth sums on the RHS vanish. The first and third sums contain N and $N(N-1)$ terms respectively, and so

$$E\left[\left(\sum_i x_i^2 \right) \left(\sum_i x_i \right)^2 \right] = N\mu_4 + N(N-1)\mu_2^2.$$

Finally, we consider the third term on the RHS of (31.45), and write

$$\left(\sum_i x_i \right)^4 = \sum_i x_i^4 + \sum_{\substack{i,j \\ j \neq i}} x_i^3 x_j + \sum_{\substack{i,j \\ j \neq i}} x_i^2 x_j^2 + \sum_{\substack{i,j,k \\ k \neq j \neq i}} x_i^2 x_j x_k + \sum_{\substack{i,j,k,l \\ l \neq k \neq j \neq i}} x_i x_j x_k x_l.$$

The expectation values of the second, fourth and fifth sums are zero, and the first and third sums contain N and $3N(N-1)$ terms respectively (for the third sum, there are $N(N-1)/2$ ways of choosing i and j, and the multinomial coefficient of $x_i^2 x_j^2$ is $4!/(2!2!) = 6$). Thus

$$E\left[\left(\sum_i x_i \right)^4 \right] = N\mu_4 + 3N(N-1)\mu_2^2.$$

Collecting together terms, we therefore obtain

$$E[s^4] = \frac{(N-1)^2}{N^3}\mu_4 + \frac{(N-1)(N^2-2N+3)}{N^3}\mu_2^2, \quad (31.46)$$

which, together with the result (31.43), may be substituted into (31.44) to obtain finally

$$V[s^2] = \frac{(N-1)^2}{N^3}\mu_4 - \frac{(N-1)(N-3)}{N^3}\mu_2^2$$
$$= \frac{N-1}{N^3}[(N-1)\nu_4 - (N-3)\nu_2^2], \quad (31.47)$$

where in the last line we have used again the fact that, since the population mean is zero, $\mu_r = \nu_r$. However, result (31.47) holds even when the population mean is not zero. ◄

From (31.43), we see that s^2 is a *biased* estimator of σ^2, although the bias becomes negligible for large N. However, it immediately follows that an unbiased estimator of σ^2 is given simply by

$$\widehat{\sigma^2} = \frac{N}{N-1}s^2, \tag{31.48}$$

where the multiplicative factor $N/(N-1)$ is often called *Bessel's correction*. Thus in terms of the sample values x_i, $i = 1, 2, \ldots, N$, an unbiased estimator of the population variance σ^2 is given by

$$\widehat{\sigma^2} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2. \tag{31.49}$$

Using (31.47), we find that the variance of the estimator $\widehat{\sigma^2}$ is

$$V[\widehat{\sigma^2}] = \left(\frac{N}{N-1}\right)^2 V[s^2] = \frac{1}{N}\left(\nu_4 - \frac{N-3}{N-1}\nu_2^2\right),$$

where ν_r is the rth central moment of the parent population. We note that, since $E[\widehat{\sigma^2}] = \sigma^2$ and $V[\widehat{\sigma^2}] \to 0$ as $N \to \infty$, the statistic $\widehat{\sigma^2}$ is also a consistent estimator of the population variance.

31.4.3 Population standard deviation σ

The standard deviation σ of a population is defined as the positive square root of the population variance σ^2 (as, indeed, our notation suggests). Thus, it is common practice to take the positive square root of the variance estimator as our estimator for σ. Thus, we take

$$\hat{\sigma} = \left(\widehat{\sigma^2}\right)^{1/2}, \tag{31.50}$$

where $\widehat{\sigma^2}$ is given by either (31.41) or (31.48), depending on whether the population mean μ is known or unknown. Because of the square root in the definition of $\hat{\sigma}$, it is not possible in either case to obtain an exact expression for $E[\hat{\sigma}]$ and $V[\hat{\sigma}]$. Indeed, although in each case the estimator is the positive square root of an unbiased estimator of σ^2, it is *not* itself an unbiased estimator of σ. However, the bias does becomes negligible for large N.

►*Obtain approximate expressions for $E[\hat{\sigma}]$ and $V[\hat{\sigma}]$ for a sample of size N in the case where the population mean μ is unknown.*

As the population mean is unknown, we use (31.50) and (31.48) to write our estimator in the form

$$\hat{\sigma} = \left(\frac{N}{N-1}\right)^{1/2} s,$$

where s is the sample standard deviation. The expectation value of this estimator is given by

$$E[\hat{\sigma}] = \left(\frac{N}{N-1}\right)^{1/2} E[(s^2)^{1/2}] \approx \left(\frac{N}{N-1}\right)^{1/2} (E[s^2])^{1/2} = \sigma.$$

An approximate expression for the variance of $\hat{\sigma}$ may be found using (31.47) and is given by

$$\begin{aligned} V[\hat{\sigma}] = \frac{N}{N-1} V[(s^2)^{1/2}] &\approx \frac{N}{N-1}\left[\frac{d}{d(s^2)}(s^2)^{1/2}\right]^2_{s^2=E[s^2]} V[s^2] \\ &\approx \frac{N}{N-1}\left[\frac{1}{4s^2}\right]_{s^2=E[s^2]} V[s^2]. \end{aligned}$$

Using the expressions (31.43) and (31.47) for $E[s^2]$ and $V[s^2]$ respectively, we obtain

$$V[\hat{\sigma}] \approx \frac{1}{4N\nu_2}\left(\nu_4 - \frac{N-3}{N-1}\nu_2^2\right). \blacktriangleleft$$

31.4.4 Population moments μ_r

We may straightforwardly generalise our discussion of estimation of the population mean $\mu\,(=\mu_1)$ in subsection 31.4.1 to the estimation of the rth population moment μ_r. An obvious choice of estimator is the rth sample moment m_r. The expectation value of m_r is given by

$$E[m_r] = \frac{1}{N}\sum_{i=1}^{N} E[x_i^r] = \frac{N\mu_r}{N} = \mu_r,$$

and so it is an unbiased estimator of μ_r.

The variance of m_r may be found in a similar manner, although the calculation is a little more complicated. We find that

$$\begin{aligned} V[m_r] &= E[(m_r - \mu_r)^2] \\ &= \frac{1}{N^2}E\left[\left(\sum_i x_i^r - N\mu_r\right)^2\right] \\ &= \frac{1}{N^2}E\left[\sum_i x_i^{2r} + \sum_i\sum_{j\neq i} x_i^r x_j^r - 2N\mu_r\sum_i x_i^r + N^2\mu_r^2\right] \\ &= \frac{1}{N}\mu_{2r} - \mu_r^2 + \frac{1}{N^2}\sum_i\sum_{j\neq i} E[x_i^r x_j^r]. \end{aligned} \qquad (31.51)$$

However, since the sample values x_i are assumed to be independent, we have

$$E[x_i^r x_j^r] = E[x_i^r]E[x_j^r] = \mu_r^2. \tag{31.52}$$

The number of terms in the sum on the RHS of (31.51) is $N(N-1)$, and so we find

$$V[m_r] = \frac{1}{N}\mu_{2r} - \mu_r^2 + \frac{N-1}{N}\mu_r^2 = \frac{\mu_{2r} - \mu_r^2}{N}. \tag{31.53}$$

Since $E[m_r] = \mu_r$ and $V[m_r] \to 0$ as $N \to \infty$, the rth sample moment m_r is also a consistent estimator of μ_r.

►*Find the covariance of the sample moments m_r and m_s for a sample of size N.*

We obtain the covariance of the sample moments m_r and m_s in a similar manner to that used above to obtain the variance of m_r. From the definition of covariance, we have

$$\begin{aligned}
\mathrm{Cov}[m_r, m_s] &= E[(m_r - \mu_r)(m_s - \mu_s)] \\
&= \frac{1}{N^2}E\left[\left(\sum_i x_i^r - N\mu_r\right)\left(\sum_j x_j^s - N\mu_s\right)\right] \\
&= \frac{1}{N^2}E\left[\sum_i x_i^{r+s} + \sum_i \sum_{j\neq i} x_i^r x_j^s - N\mu_r \sum_j x_j^s - N\mu_s \sum_i x_i^r + N^2\mu_r\mu_s\right]
\end{aligned}$$

Assuming the x_i to be independent, we may again use result (31.52) to obtain

$$\begin{aligned}
\mathrm{Cov}[m_r, m_s] &= \frac{1}{N^2}[N\mu_{r+s} + N(N-1)\mu_r\mu_s - N^2\mu_r\mu_s - N^2\mu_s\mu_r + N^2\mu_r\mu_s] \\
&= \frac{1}{N}\mu_{r+s} + \frac{N-1}{N}\mu_r\mu_s - \mu_r\mu_s \\
&= \frac{\mu_{r+s} - \mu_r\mu_s}{N}.
\end{aligned}$$

We note that by setting $r = s$, we recover the expression (31.53) for $V[m_r]$. ◄

31.4.5 Population central moments ν_r

We may generalise the discussion of estimators for the second central moment ν_2 (or equivalently σ^2) given in subsection 31.4.2 to the estimation of the rth central moment ν_r. In particular, we saw in that subsection that our choice of estimator for ν_2 depended on whether the population mean μ_1 is known; the same is true for the estimation of ν_r.

Let us first consider the case in which μ_1 is known. From (30.54), we may write ν_r as

$$\nu_r = \mu_r - {}^rC_1\mu_{r-1}\mu_1 + \cdots + (-1)^k\, {}^rC_k\mu_{r-k}\mu_1^k + \cdots + (-1)^{r-1}({}^rC_{r-1} - 1)\mu_1^r.$$

If μ_1 is known, a suitable estimator is obviously

$$\hat{\nu}_r = m_r - {}^rC_1 m_{r-1}\mu_1 + \cdots + (-1)^k\, {}^rC_k m_{r-k}\mu_1^k + \cdots + (-1)^{r-1}({}^rC_{r-1} - 1)\mu_1^r,$$

where m_r is the rth sample moment. Since μ_1 and the binomial coefficients are (known) constants, it is immediately clear that $E[\hat{\nu}_r] = \nu_r$, and so $\hat{\nu}_r$ is an unbiased estimator of ν_r. It is also possible to obtain an expression for $V[\hat{\nu}_r]$, though the calculation is somewhat lengthy.

In the case where the population mean μ_1 is *not* known, the situation is more complicated. We saw in subsection 31.4.2 that the second sample moment n_2 (or s^2) is *not* an unbiased estimator of ν_2 (or σ^2). Similarly, the rth *central* moment of a sample, n_r, is not an unbiased estimator of the rth population central moment ν_r. However, in all cases the bias becomes negligible in the limit of large N.

As we also found in the same subsection, there are complications in calculating the expectation and variance of n_2; these complications increase considerably for general r. Nevertheless, we have derived already in this chapter *exact* expressions for the expectation value of the first few sample central moments, which are valid for samples of any size N. From (31.40), (31.43) and (31.46), we find

$$\begin{aligned} E[n_1] &= 0, \\ E[n_2] &= \frac{N-1}{N}\nu_2, \\ E[n_2^2] &= \frac{N-1}{N^3}[(N-1)\nu_4 + (N^2-2N+3)\nu_2^2]. \end{aligned} \tag{31.54}$$

By similar arguments it can be shown that

$$E[n_3] = \frac{(N-1)(N-2)}{N^2}\nu_3, \tag{31.55}$$

$$E[n_4] = \frac{N-1}{N^3}[(N^2-3N+3)\nu_4 + 3(2N-3)\nu_2^2]. \tag{31.56}$$

From (31.54) and (31.55), we see that unbiased estimators of ν_2 and ν_3 are

$$\hat{\nu}_2 = \frac{N}{N-1}n_2, \tag{31.57}$$

$$\hat{\nu}_3 = \frac{N^2}{(N-1)(N-2)}n_3, \tag{31.58}$$

where (31.57) simply re-establishes our earlier result that $\widehat{\sigma^2} = Ns^2/(N-1)$ is an unbiased estimator of σ^2.

Unfortunately, the pattern that appears to be emerging in (31.57) and (31.58) is *not* continued for higher r, as is seen immediately from (31.56). Nevertheless, in the limit of large N, the bias becomes negligible, and often one simply takes $\hat{\nu}_r = n_r$. For large N, it may be shown that

$$\begin{aligned} E[n_r] &\approx \nu_r \\ V[n_r] &\approx \frac{1}{N}(\nu_{2r} - \nu_r^2 + r^2\nu_2\nu_{r-1}^2 - 2r\nu_{r-1}\nu_{r+1}) \\ \mathrm{Cov}[n_r, n_s] &\approx \frac{1}{N}(\nu_{r+s} - \nu_r\nu_s + rs\nu_2\nu_{r-1}\nu_{s-1} - r\nu_{r-1}\nu_{s+1} - s\nu_{s-1}\nu_{r+1}) \end{aligned}$$

31.4.6 Population covariance Cov[x, y] and correlation Corr[x, y]

So far we have assumed that each of our N independent samples consists of a single number x_i. Let us now extend our discussion to a situation in which each sample consists of two numbers x_i, y_i, which we may consider as being drawn randomly from a two-dimensional population $P(x, y)$. In particular, we now consider estimators for the population covariance $\mathrm{Cov}[x, y]$ and for the correlation $\mathrm{Corr}[x, y]$.

When μ_x and μ_y are *known*, an appropriate estimator of the population covariance is

$$\widehat{\mathrm{Cov}}[x, y] = \overline{xy} - \mu_x\mu_y = \left(\frac{1}{N}\sum_{i=1}^{N} x_i y_i\right) - \mu_x\mu_y. \qquad (31.59)$$

This estimator is unbiased since

$$E\left[\widehat{\mathrm{Cov}}[x, y]\right] = \frac{1}{N}E\left[\sum_{i=1}^{N} x_i y_i\right] - \mu_x\mu_y = E[x_i y_i] - \mu_x\mu_y = \mathrm{Cov}[x, y].$$

Alternatively, if μ_x and μ_y are *unknown*, it is natural to replace μ_x and μ_y in (31.59) by the sample means $\bar{x}$ and $\bar{y}$ respectively, in which case we recover the sample covariance $V_{xy} = \overline{xy} - \bar{x}\bar{y}$ discussed in subsection 31.2.4. This estimator is biased but an unbiased estimator of the population covariance is obtained by forming

$$\widehat{\mathrm{Cov}}[x, y] = \frac{N}{N-1}V_{xy}. \qquad (31.60)$$

►*Calculate the expectation value of the sample covariance V_{xy} for a sample of size N.*

The sample covariance is given by

$$V_{xy} = \left(\frac{1}{N}\sum_i x_i y_i\right) - \left(\frac{1}{N}\sum_i x_i\right)\left(\frac{1}{N}\sum_j y_j\right).$$

Thus its expectation value is given by

$$\begin{aligned} E[V_{xy}] &= \frac{1}{N}E\left[\sum_i x_i y_i\right] - \frac{1}{N^2}E\left[\left(\sum_i x_i\right)\left(\sum_j x_j\right)\right] \\ &= E[x_i y_i] - \frac{1}{N^2}E\left[\sum_i x_i y_i + \sum_{\substack{i,j \\ j\neq i}} x_i y_j\right] \end{aligned}$$

Since the number of terms in the double sum on the RHS is $N(N-1)$, we have

$$\begin{aligned}
E[V_{xy}] &= E[x_iy_i] - \frac{1}{N^2}(NE[x_iy_i] + N(N-1)E[x_iy_j]) \\
&= E[x_iy_i] - \frac{1}{N^2}(NE[x_iy_i] + N(N-1)E[x_i]E[y_j]) \\
&= E[x_iy_i] - \frac{1}{N}\left(E[x_iy_i] + (N-1)\mu_x\mu_y\right) = \frac{N-1}{N}\mathrm{Cov}[x,y],
\end{aligned}$$

where we have used the fact that, since the samples are independent, $E[x_iy_j] = E[x_i]E[y_j]$. ◀

It is possible to obtain expressions for the variances of the estimators (31.59) and (31.60) but these quantities depend upon higher moments of the population $P(x, y)$ and are extremely lengthy to calculate.

Whether the means μ_x and μ_y are known or unknown, an estimator of the population correlation $\mathrm{Corr}[x, y]$ is given by

$$\widehat{\mathrm{Corr}}[x,y] = \frac{\widehat{\mathrm{Cov}}[x,y]}{\hat{\sigma}_x\hat{\sigma}_y}, \tag{31.61}$$

where $\widehat{\mathrm{Cov}}[x, y]$, $\hat{\sigma}_x$ and $\hat{\sigma}_y$ are the appropriate estimators of the population covariance and standard deviations. Although this estimator is only asymptotically unbiased, i.e. for large N, it is widely used because of its simplicity. Once again the variance of the estimator depends on the higher moments of $P(x, y)$ and is difficult to calculate.

In the case in which the means μ_x and μ_y are unknown, a suitable (but biased) estimator is

$$\widehat{\mathrm{Corr}}[x,y] = \frac{N}{N-1}\frac{V_{xy}}{s_xs_y} = \frac{N}{N-1}r_{xy}, \tag{31.62}$$

where s_x and s_y are the sample standard deviations of the x_i and y_i respectively and r_{xy} is the sample correlation. In the special case when the parent population $P(x, y)$ is Gaussian, it may be shown that, if $\rho = \mathrm{Corr}[x, y]$,

$$E[r_{xy}] = \rho - \frac{\rho(1-\rho^2)}{2N} + \mathrm{O}(N^{-2}), \tag{31.63}$$

$$V[r_{xy}] = \frac{1}{N}(1-\rho^2)^2 + \mathrm{O}(N^{-2}), \tag{31.64}$$

from which the expectation value and variance of the estimator $\widehat{\mathrm{Corr}}[x, y]$ may be found immediately.

We note finally that our discussion may be extended, without significant alteration, to the general case in which each data item consists of n numbers $x_i, y_i, \ldots, z_i$.

31.4.7 A worked example

To conclude our discussion of basic estimators, we reconsider the set of experimental data given in subsection 31.2.4. We carry the analysis as far as calculating the standard errors in the estimated population parameters, including the population correlation.

►*Ten UK citizens are selected at random and their heights and weights are found to be as follows (to the nearest* cm *or* kg *respectively):*

Person	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Height (cm)	194	168	177	180	171	190	151	169	175	182
Weight (kg)	75	53	72	80	75	75	57	67	46	68

Estimate the means, μ_x and μ_y, and standard deviations, σ_x and σ_y, of the two-dimensional joint population from which the sample was drawn, quoting the standard error on the estimate in each case. Estimate also the correlation $\text{Corr}[x, y]$ *of the population, and quote the standard error on the estimate under the assumption that the population is a multivariate Gaussian.*

In subsection 31.2.4, we calculated various sample statistics for these data. In particular, we found that for our sample of size $N = 10$,

$$\bar{x} = 175.7, \qquad \bar{y} = 66.8,$$

$$s_x = 11.6, \qquad s_y = 10.6, \qquad r_{xy} = 0.54.$$

Let us begin by estimating the means μ_x and μ_y. As discussed in subsection 31.4.1, the sample mean is an unbiased, consistent estimator of the population mean. Moreover, the standard error on $\bar{x}$ (say) is $\sigma_x/\sqrt{N}$. In this case, however, we do not know the true value of σ_x and we must estimate it using $\widehat{\sigma}_x = \sqrt{N/(N-1)}s_x$. Thus, our estimates of μ_x and μ_y, with associated standard errors, are

$$\hat{\mu}_x = \bar{x} \pm \frac{s_x}{\sqrt{N-1}} = 175.7 \pm 3.9,$$
$$\hat{\mu}_y = \bar{y} \pm \frac{s_y}{\sqrt{N-1}} = 66.8 \pm 3.5.$$

We now turn to estimating σ_x and σ_y. As just mentioned, our estimate of σ_x (say) is $\widehat{\sigma}_x = \sqrt{N/(N-1)}s_x$. Its variance (see the final line of subsection 31.4.3) is given approximately by

$$V[\hat{\sigma}] \approx \frac{1}{4N\nu_2}\left(\nu_4 - \frac{N-3}{N-1}\nu_2^2\right).$$

Since we do not know the true values of the population central moments ν_2 and ν_4, we must use their estimated values in this expression. We may take $\hat{\nu}_2 = \widehat{\sigma_x^2} = (\hat{\sigma})^2$, which we have already calculated. It still remains, however, to estimate ν_4. As implied near the end of subsection 31.4.5, it is acceptable to take $\hat{\nu}_4 = n_4$. Thus for the x_i and y_i values, we have

$$(\hat{\nu}_4)_x = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^4 = 53\,411.6$$

$$(\hat{\nu}_4)_y = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^4 = 27\,732.5$$

Substituting these values into (31.50), we obtain

$$\hat{\sigma}_x = \left(\frac{N}{N-1}\right)^{1/2} s_x \pm (\hat{V}[\hat{\sigma}_x])^{1/2} = 12.2 \pm 6.7, \tag{31.65}$$

$$\hat{\sigma}_y = \left(\frac{N}{N-1}\right)^{1/2} s_y \pm (\hat{V}[\hat{\sigma}_y])^{1/2} = 11.2 \pm 3.6. \tag{31.66}$$

Finally, we estimate the population correlation Corr$[x, y]$, which we shall denote by ρ. From (31.62), we have

$$\hat{\rho} = \frac{N}{N-1} r_{xy} = 0.60.$$

Under the *assumption* that the sample was drawn from a two-dimensional Gaussian population $P(x, y)$, the variance of our estimator is given by (31.64). Since we do not know the true value of ρ, we must use our estimate $\hat{\rho}$. Thus, we find that the standard error $\Delta\rho$ in our estimate is given approximately by

$$\Delta\rho \approx \frac{10}{9}\left(\frac{1}{10}\right)[1 - (0.60)^2]^2 = 0.05. \blacktriangleleft$$

31.5 Maximum-likelihood method

The population from which the sample $x_1, x_2, \ldots, x_N$ is drawn is, in general, *unknown*. In the previous section, we assumed that the sample values were independent and drawn from a one-dimensional population $P(x)$, and we considered basic estimators of the moments and central moments of $P(x)$. We did *not*, however, assume a particular functional form for $P(x)$. We now discuss the process of *data modelling*, in which a specific form is assumed for the population.

In the most general case, it will not be known whether the sample values are independent, and so let us consider the full joint population $P(\mathbf{x})$, where $\mathbf{x}$ is the point in the N-dimensional data space with coordinates $x_1, x_2, \ldots, x_N$. We then adopt the *hypothesis* H that the probability distribution of the sample values has some particular functional form $L(\mathbf{x}; \mathbf{a})$, dependent on the values of some set of *parameters* a_i, $i = 1, 2, \ldots, m$. Thus, we have

$$P(\mathbf{x}|\mathbf{a}, H) = L(\mathbf{x}; \mathbf{a}),$$

where we make explicit the conditioning on both the assumed functional form and on the parameter values. $L(\mathbf{x}; \mathbf{a})$ is called the *likelihood function*. Hypotheses of this type form the basis of *data modelling* and *parameter estimation*. One proposes a particular model for the underlying population and then attempts to estimate from the sample values $x_1, x_2, \ldots, x_N$ the values of the parameters $\mathbf{a}$ defining this model.

►*A company measures the duration (in minutes) of the N intervals x_i, $i = 1, 2, \ldots, N$ between successive telephone calls received by its switchboard. Suppose that the sample values x_i are drawn independently from the distribution $P(x|\tau) = (1/\tau)\exp(-x/\tau)$, where τ is the mean interval between calls. Calculate the likelihood function $L(\mathbf{x}; \tau)$.*

Since the sample values are independent and drawn from the stated distribution, the

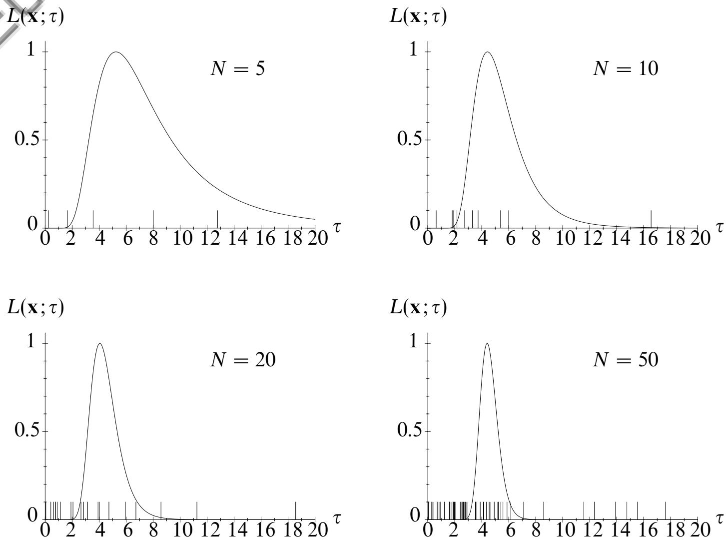


Figure 31.5 Examples of the likelihood function (31.67) for samples of different size N. In each case, the true value of the parameter is $\tau = 4$ and the sample values x_i are indicated by the short vertical lines. For the purposes of illustration, in each case the likelihood function is normalised so that its maximum value is unity.

likelihood is given by

$$\begin{aligned} L(\mathbf{x};\tau) &= P(x_i|\tau)P(x_2|\tau)\cdots P(x_N|\tau) \\ &= \frac{1}{\tau}\exp\left(-\frac{x_1}{\tau}\right)\frac{1}{\tau}\exp\left(-\frac{x_2}{\tau}\right)\cdots\frac{1}{\tau}\exp\left(-\frac{x_N}{\tau}\right) \\ &= \frac{1}{\tau^N}\exp\left[-\frac{1}{\tau}(x_1+x_2+\cdots+x_N)\right]. \end{aligned} \qquad (31.67)$$

which is to be considered as a function of τ, given that the sample values x_i are fixed. ◄

The likelihood function (31.67) depends on just a single parameter τ. Plots of the likelihood function, considered as a function of τ, are shown in figure 31.5 for samples of different size N. The true value of the parameter τ used to generate the sample values was 4. In each case, the sample values x_i are indicated by the short vertical lines. For the purposes of illustration, the likelihood function in each case has been scaled so that its maximum value is unity (this is, in fact, common practice). We see that when the sample size is small, the likelihood function is very broad. As N increases, however, the function becomes narrower (its width is inversely proportional to $\sqrt{N}$) and tends to a Gaussian-like shape, with its peak centred on 4, the true value of τ. We discuss these properties of the likelihood function in more detail in subsection 31.5.6.

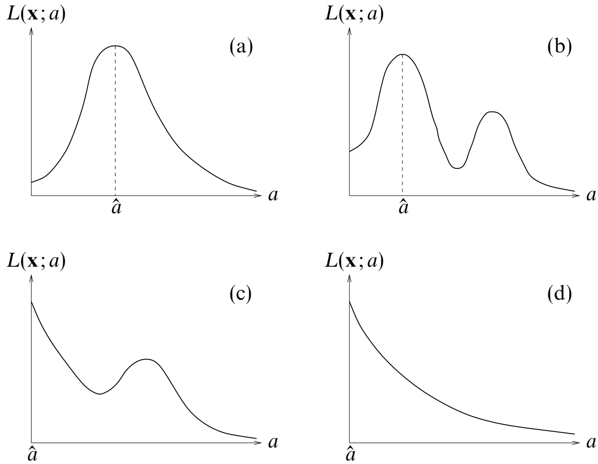


Figure 31.6 Typical shapes of one-dimensional likelihood functions $L(\mathbf{x}; a)$ encountered in practice, when, for illustration purposes, it is assumed that the parameter a is restricted to the range zero to infinity. The ML estimator in the various cases occurs at: (a) the only stationary point; (b) one of several stationary points; (c) an end-point of the allowed parameter range that is not a stationary point (although stationary points do exist); (d) an end-point of the allowed parameter range in which no stationary point exists.

31.5.1 The maximum-likelihood estimator

Since the likelihood function $L(\mathbf{x}; \mathbf{a})$ gives the probability density associated with any particular set of values of the parameters $\mathbf{a}$, our best estimate $\hat{\mathbf{a}}$ of these parameters is given by the values of $\mathbf{a}$ for which $L(\mathbf{x}; \mathbf{a})$ is a maximum. This is called the *maximum-likelihood estimator* (or ML estimator).

In general, the likelihood function can have a complicated shape when considered as a function of $\mathbf{a}$, particularly when the dimensionality of the space of parameters $a_1, a_2, \ldots, a_M$ is large. It may be that the values of some parameters are either known or assumed in advance, in which case the effective dimensionality of the likelihood function is reduced accordingly. However, even when the likelihood depends on just a single parameter a (either intrinsically or as the result of assuming particular values for the remaining parameters), its form may be complicated when the sample size N is small. Frequently occurring shapes of one-dimensional likelihood functions are illustrated in figure 31.6, where we have assumed, for definiteness, that the allowed range of the parameter a is zero to infinity. In each case, the ML estimate $\hat{a}$ is also indicated. Of course, the 'shape' of higher-dimensional likelihood functions may be considerably more complicated.

In many simple cases, however, the likelihood function $L(\mathbf{x}; \mathbf{a})$ has a single maximum that occurs at a stationary point (the likelihood function is then termed *unimodal*). In this case, the ML estimators of the parameters a_i, $i = 1, 2, \ldots, M$, may be found *without* evaluating the full likelihood function $L(\mathbf{x}; \mathbf{a})$. Instead, one simply solves the M simultaneous equations

$$\left.\frac{\partial L}{\partial a_i}\right|_{\mathbf{a}=\hat{\mathbf{a}}} = 0 \qquad \text{for } i = 1, 2, \ldots, M. \tag{31.68}$$

Since $\ln z$ is a monotonically increasing function of z (and therefore has the same stationary points), it is often more convenient, in fact, to maximise the *log-likelihood function*, $\ln L(\mathbf{x}; \mathbf{a})$, with respect to the a_i. Thus, one may, as an alternative, solve the equations

$$\left.\frac{\partial \ln L}{\partial a_i}\right|_{\mathbf{a}=\hat{\mathbf{a}}} = 0 \qquad \text{for } i = 1, 2, \ldots, M. \tag{31.69}$$

Clearly, (31.68) and (31.69) will lead to the same ML estimates $\hat{\mathbf{a}}$ of the parameters. In either case, it is, of course, prudent to check that the point $\mathbf{a} = \hat{\mathbf{a}}$ is a local maximum.

►*Find the ML estimate of the parameter τ in the previous example, in terms of the measured values x_i, $i = 1, 2, \ldots, N$.*

From (31.67), the log-likelihood function in this case is given by

$$\ln L(\mathbf{x}; \tau) = \sum_{i=1}^{N} \ln\left(\frac{1}{\tau} e^{-x_i/\tau}\right) = -\sum_{i=1}^{N}\left(\ln \tau + \frac{x_i}{\tau}\right). \tag{31.70}$$

Differentiating with respect to the parameter τ and setting the result equal to zero, we find

$$\frac{\partial \ln L}{\partial \tau} = -\sum_{i=1}^{N}\left(\frac{1}{\tau} - \frac{x_i}{\tau^2}\right) = 0.$$

Thus the ML estimate of the parameter τ is given by

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{31.71}$$

which is simply the sample mean of the N measured intervals. ◄

In the previous example we assumed that the sample values x_i were drawn independently from the *same* parent distribution. The ML method is more flexible than this restriction might seem to imply, and it can equally well be applied to the common case in which the samples x_i are independent but each is drawn from a *different* distribution.

►*In an experiment, N independent measurements x_i of some quantity are made. Suppose that the random measurement error on the ith sample value is Gaussian distributed with mean zero and known standard deviation σ_i. Calculate the ML estimate of the true value μ of the quantity being measured.*

As the measurements are independent, the likelihood factorises:

$$L(\mathbf{x};\mu,\{\sigma_k\}) = \prod_{i=1}^{N} P(x_i|\mu,\sigma_i),$$

where $\{\sigma_k\}$ denotes collectively the set of known standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_N$. The individual distributions are given by

$$P(x_i|\mu,\sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma_i^2}\right].$$

and so the full log-likelihood function is given by

$$\ln L(\mathbf{x};\mu,\{\sigma_k\}) = -\frac{1}{2}\sum_{i=1}^{N}\left[\ln(2\pi\sigma_i^2) + \frac{(x_i-\mu)^2}{\sigma_i^2}\right].$$

Differentiating this expression with respect to μ and setting the result equal to zero, we find

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{N} \frac{x_i-\mu}{\sigma_i^2} = 0,$$

from which we obtain the ML estimator

$$\hat{\mu} = \frac{\sum_{i=1}^{N}(x_i/\sigma_i^2)}{\sum_{i=1}^{N}(1/\sigma_i^2)}. \tag{31.72}$$

This estimator is commonly used when averaging data with different *statistical weights* $w_i = 1/\sigma_i^2$. We note that when all the variances σ_i^2 have the same value the estimator reduces to the sample mean of the data x_i. ◄

There is, in fact, no requirement in the ML method that the sample values be independent. As an illustration, we shall generalise the above example to a case in which the measurements x_i are not all independent. This would occur, for example, if these measurements were based at least in part on the same data.

►*In an experiment N measurements x_i of some quantity are made. Suppose that the random measurement errors on the samples are drawn from a joint Gaussian distribution with mean zero and known covariance matrix V. Calculate the ML estimate of the true value μ of the quantity being measured.*

From (30.148), the likelihood in this case is given by

$$L(\mathbf{x};\mu,\mathsf{V}) = \frac{1}{(2\pi)^{N/2}|\mathsf{V}|^{1/2}} \exp\left[-\tfrac{1}{2}(\mathsf{x}-\mu\mathsf{1})^{\mathrm{T}}\mathsf{V}^{-1}(\mathsf{x}-\mu\mathsf{1})\right],$$

where x is the column matrix with components $x_1, x_2, \ldots, x_N$ and $\mathsf{1}$ is the column matrix with all components equal to unity. Thus, the log-likelihood function is given by

$$\ln L(\mathbf{x};\mu,\mathsf{V}) = -\tfrac{1}{2}\left[N\ln(2\pi) + \ln|\mathsf{V}| + (\mathsf{x}-\mu\mathsf{1})^{\mathrm{T}}\mathsf{V}^{-1}(\mathsf{x}-\mu\mathsf{1})\right].$$

Differentiating with respect to μ and setting the result equal to zero gives

$$\frac{\partial \ln L}{\partial \mu} = \mathsf{1}^{\mathrm{T}}\mathsf{V}^{-1}(\mathsf{x} - \mu\mathsf{1}) = 0.$$

Thus, the ML estimator is given by

$$\hat{\mu} = \frac{\mathsf{1}^{\mathrm{T}}\mathsf{V}^{-1}\mathsf{x}}{\mathsf{1}^{\mathrm{T}}\mathsf{V}^{-1}\mathsf{1}} = \frac{\sum_{i,j}(V^{-1})_{ij}x_j}{\sum_{i,j}(V^{-1})_{ij}}.$$

In the case of uncorrelated errors in measurement, $(V^{-1})_{ij} = \delta_{ij}/\sigma_i^2$ and our estimator reduces to that given in (31.72). ◄

In all the examples considered so far, the likelihood function has been effectively one-dimensional, either instrinsically or under the assumption that the values of all but one of the parameters are known in advance. As the following example involving two parameters shows, the application of the ML method to the estimation of several parameters simultaneously is straightforward.

►*In an experiment N measurements x_i of some quantity are made. Suppose the random error on each sample value is drawn independently from a Gaussian distribution of mean zero but unknown standard deviation σ (which is the same for each measurement). Calculate the ML estimates of the true value μ of the quantity being measured and the standard deviation σ of the random errors.*

In this case the log-likelihood function is given by

$$\ln L(\mathbf{x}; \mu, \sigma) = -\frac{1}{2}\sum_{i=1}^{N}\left[\ln(2\pi\sigma^2) + \frac{(x_i - \mu)^2}{\sigma^2}\right].$$

Taking partial derivatives of $\ln L$ with respect to μ and σ and setting the results equal to zero at the joint estimate $\hat{\mu}, \hat{\sigma}$, we obtain

$$\sum_{i=1}^{N} \frac{x_i - \hat{\mu}}{\hat{\sigma}^2} = 0, \qquad (31.73)$$

$$\sum_{i=1}^{N} \frac{(x_i - \hat{\mu})^2}{\hat{\sigma}^3} - \sum_{i=1}^{N} \frac{1}{\hat{\sigma}} = 0. \qquad (31.74)$$

In principle, one should solve these two equations simultaneously for $\hat{\mu}$ and $\hat{\sigma}$, but in this case we notice that the first is solved immediately by

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i = \bar{x},$$

where $\bar{x}$ is the sample mean. Substituting this result into the second equation, we find

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} = s,$$

where s is the sample standard deviation. As shown in subsection 31.4.3, s is a biased estimator of σ. The reason why the ML method may produce a biased estimator is discussed in the next subsection. ◄

31.5.2 Transformation invariance and bias of ML estimators

An extremely useful property of ML estimators is that they are *invariant* to parameter transformations. Suppose that, instead of estimating some parameter a of the assumed population, we wish to estimate some function $\alpha(a)$ of the parameter. The ML estimator $\hat{\alpha}(a)$ is given by the value assumed by the function $\alpha(a)$ at the maximum point of the likelihood, which is simply equal to $\alpha(\hat{a})$. Thus, we have the very convenient property

$$\hat{\alpha}(a) = \alpha(\hat{a}).$$

We do not have to worry about the distinction between estimating a and estimating a function of a. This is *not* true, in general, for other estimation procedures.

►*A company measures the duration (in minutes) of the N intervals x_i, $i = 1, 2, \ldots, N$, between successive telephone calls received by its switchboard. Suppose that the sample values x_i are drawn independently from the distribution $P(x|\tau) = (1/\tau)\exp(-x/\tau)$. Find the ML estimate of the parameter $\lambda = 1/\tau$.*

This is the same problem as the first one considered in subsection 31.5.1. In terms of the new parameter λ, the log-likelihood function is given by

$$\ln L(\mathbf{x}; \lambda) = \sum_{i=1}^{N} \ln(\lambda e^{-\lambda x_i}) = \sum_{i=1}^{N} (\ln \lambda - \lambda x_i).$$

Differentiating with respect to λ and setting the result equal to zero, we have

$$\frac{\partial \ln L}{\partial \lambda} = \sum_{i=1}^{N} \left(\frac{1}{\lambda} - x_i \right) = 0.$$

Thus, the ML estimator of the parameter λ is given by

$$\hat{\lambda} = \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right)^{-1} = \bar{x}^{-1}. \qquad (31.75)$$

Referring back to (31.71), we see that, as expected, the ML estimators of λ and τ are related by $\hat{\lambda} = 1/\hat{\tau}$. ◄

Although this invariance property is useful it also means that, in general, ML estimators may be *biased*. In particular, one must be aware of the fact that even if $\hat{a}$ is an unbiased ML estimator of a it does *not* follow that the estimator $\hat{\alpha}(a)$ is also unbiased. In the limit of large N, however, the bias of ML estimators always tends to zero. As an illustration, it is straightforward to show (see exercise 31.8) that the ML estimators $\hat{\tau}$ and $\hat{\lambda}$ in the above example have expectation values

$$E[\hat{\tau}] = \tau \qquad \text{and} \qquad E[\hat{\lambda}] = \frac{N}{N-1}\lambda. \qquad (31.76)$$

In fact, since $\hat{\tau} = \bar{x}$ and the sample values are independent, the first result follows immediately from (31.40). Thus, $\hat{\tau}$ is unbiased, but $\hat{\lambda} = 1/\hat{\tau}$ is biased, albeit that the bias tends to zero for large N.

31.5.3 Efficiency of ML estimators

We showed in subsection 31.3.2 that Fisher's inequality puts a lower limit on the variance $V[\hat{a}]$ of any estimator of the parameter a. Under our hypothesis H on p. 1255, the functional form of the population is given by the likelihood function, i.e. $P(\mathbf{x}|\mathbf{a}, H) = L(\mathbf{x};\mathbf{a})$. Thus, if this hypothesis is correct, we may replace P by L in Fisher's inequality (31.18), which then reads

$$V[\hat{a}] \geq \left(1 + \frac{\partial b}{\partial a}\right)^2 \bigg/ E\left[-\frac{\partial^2 \ln L}{\partial a^2}\right],$$

where b is the bias in the estimator $\hat{a}$. We usually denote the RHS by $V_{\min}$.

An important property of ML estimators is that *if* there exists an efficient estimator $\hat{a}_{\rm eff}$, i.e. one for which $V[\hat{a}_{\rm eff}] = V_{\min}$, then it *must* be the ML estimator or some function thereof. This is easily shown by replacing P by L in the proof of Fisher's inequality given in subsection 31.3.2. In particular, we note that the equality in (31.22) holds only if $h(\mathbf{x}) = cg(\mathbf{x})$, where c is a constant. Thus, if an efficient estimator $\hat{a}_{\rm eff}$ exists, this is equivalent to demanding that

$$\frac{\partial \ln L}{\partial a} = c[\hat{a}_{\rm eff} - \alpha(a)].$$

Now, the ML estimator $\hat{a}_{\rm ML}$ is given by

$$\left.\frac{\partial \ln L}{\partial a}\right|_{a=\hat{a}_{\rm ML}} = 0 \qquad \Rightarrow \qquad c[\hat{a}_{\rm eff} - \alpha(\hat{a}_{\rm ML})] = 0,$$

which, in turn, implies that $\hat{a}_{\rm eff}$ must be some function of $\hat{a}_{\rm ML}$.

►*Show that the ML estimator $\hat{\tau}$ given in (31.71) is an efficient estimator of the parameter τ.*

As shown in (31.70), the log-likelihood function in this case is

$$\ln L(\mathbf{x};\tau) = -\sum_{i=1}^{N}\left(\ln\tau + \frac{x_i}{\tau}\right).$$

Differentiating twice with respect to τ, we find

$$\frac{\partial^2 \ln L}{\partial \tau^2} = \sum_{i=1}^{N}\left(\frac{1}{\tau^2} - \frac{2x_i}{\tau^3}\right) = \frac{N}{\tau^2}\left(1 - \frac{2}{\tau N}\sum_{i=1}^{N} x_i\right), \tag{31.77}$$

and so the expectation value of this expression is

$$E\left[\frac{\partial^2 \ln L}{\partial \tau^2}\right] = \frac{N}{\tau^2}\left(1 - \frac{2}{\tau}E[x_i]\right) = -\frac{N}{\tau^2},$$

where we have used the fact that $E[x] = \tau$. Setting $b = 0$ in (31.18), we thus find that for *any* unbiased estimator of τ,

$$V[\hat{\tau}] \geq \frac{\tau^2}{N}.$$

From (31.76), we see that the ML estimator $\hat{\tau} = \sum_i x_i/N$ is unbiased. Moreover, using the fact that $V[x] = \tau^2$, it follows immediately from (31.40) that $V[\hat{\tau}] = \tau^2/N$. Thus $\hat{\tau}$ is a minimum-variance estimator of τ. ◄

31.5.4 Standard errors and confidence limits on ML estimators

The ML method provides a procedure for obtaining a particular set of estimators $\hat{\mathbf{a}}_{\rm ML}$ for the parameters $\mathbf{a}$ of the assumed population $P(\mathbf{x}|\mathbf{a})$. As for any other set of estimators, the associated standard errors, covariances and confidence intervals can be found as described in subsections 31.3.3 and 31.3.4.

►*A company measures the duration (in minutes) of the* 10 *intervals* x_i, $i = 1, 2, \ldots, 10$, *between successive telephone calls made to its switchboard to be as follows:*

0.43 0.24 3.03 1.93 1.16 8.65 5.33 6.06 5.62 5.22.

Supposing that the sample values are drawn independently from the probability distribution $P(x|\tau) = (1/\tau)\exp(-x/\tau)$, *find the ML estimate of the mean* τ *and quote an estimate of the standard error on your result.*

As shown in (31.71) the (unbiased) ML estimator $\hat{\tau}$ in this case is simply the sample mean $\bar{x} = 3.77$. Also, as shown in subsection 31.5.3, $\hat{\tau}$ is a minimum-variance estimator with $V[\hat{\tau}] = \tau^2/N$. Thus, the standard error in $\hat{\tau}$ is simply

$$\sigma_{\hat{\tau}} = \frac{\tau}{\sqrt{N}}. \qquad (31.78)$$

Since we do not know the true value of τ, however, we must instead quote an estimate $\hat{\sigma}_{\hat{\tau}}$ of the standard error, obtained by substituting our estimate $\hat{\tau}$ for τ in (31.78). Thus, we quote our final result as

$$\tau = \hat{\tau} \pm \frac{\hat{\tau}}{\sqrt{N}} = 3.77 \pm 1.19. \qquad (31.79)$$

For comparison, the true value used to create the sample was $\tau = 4$. ◄

For the particular problem considered in the above example, it is in fact possible to derive the full sampling distribution of the ML estimator $\hat{\tau}$ using characteristic functions, and it is given by

$$P(\hat{\tau}|\tau) = \frac{N^N}{(N-1)!}\frac{\hat{\tau}^{N-1}}{\tau^N}\exp\left(-\frac{N\hat{\tau}}{\tau}\right), \qquad (31.80)$$

where N is the size of the sample. This function is plotted in figure 31.7 for the case $\tau = 4$ and $N = 10$, which pertains to the above example. Knowledge of the analytic form of the sampling distribution allows one to place *confidence limits* on the estimate $\hat{\tau}$ obtained, as discussed in subsection 31.3.4.

►*Using the sample values in the above example, obtain the* 68% *central confidence interval on the value of* τ.

For the sample values given, our observed value of the ML estimator is $\hat{\tau}_{\rm obs} = 3.77$. Thus, from (31.28) and (31.29), the 68% central confidence interval $[\tau_-, \tau_+]$ on the value of τ is found by solving the equations

$$\int_{-\infty}^{\hat{\tau}_{\rm obs}} P(\hat{\tau}|\tau_+)\, d\hat{\tau} = 0.16,$$

$$\int_{\hat{\tau}_{\rm obs}}^{\infty} P(\hat{\tau}|\tau_-)\, d\hat{\tau} = 0.16,$$

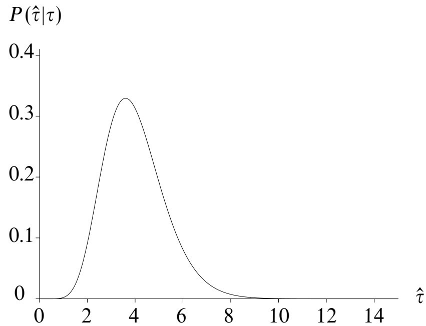


Figure 31.7 The sampling distribution $P(\hat{\tau}|\tau)$ for the estimator $\hat{\tau}$ for the case $\tau = 4$ and $N = 10$.

where $P(\hat{\tau}|\tau)$ is given by (31.80) with $N = 10$. The above integrals can be evaluated analytically but the calculations are rather cumbersome. It is much simpler to evaluate them by numerical integration, from which we find $[\tau_-, \tau_+] = [2.86, 5.46]$. Alternatively, we could quote the estimate and its 68% confidence interval as

$$\tau = 3.77\,^{+1.69}_{-0.91}.$$

Thus we see that the 68% central confidence interval is not symmetric about the estimated value, and differs from the standard error calculated above. This is a result of the (non-Gaussian) shape of the sampling distribution $P(\hat{\tau}|\tau)$, apparent in figure 31.7. ◀

In many problems, however, it is not possible to derive the full sampling distribution of an ML estimator $\hat{a}$ in order to obtain its confidence intervals. Indeed, one may not even be able to obtain an analytic formula for its standard error $\sigma_{\hat{a}}$. This is particularly true when one is estimating several parameter $\hat{\mathbf{a}}$ simultaneously, since the joint sampling distribution will be, in general, very complicated. Nevertheless, as we discuss below, the likelihood function $L(\mathbf{x}; \mathbf{a})$ *itself* can be used very simply to obtain standard errors and confidence intervals. The justification for this has its roots in the *Bayesian* approach to statistics, as opposed to the more traditional *frequentist* approach we have adopted here. We now give a brief discussion of the Bayesian viewpoint on parameter estimation.

31.5.5 The Bayesian interpretation of the likelihood function

As stated at the beginning of section 31.5, the likelihood function $L(\mathbf{x}; \mathbf{a})$ is defined by

$$P(\mathbf{x}|\mathbf{a}, H) = L(\mathbf{x}; \mathbf{a}),$$

where H denotes our hypothesis of an assumed functional form. Now, using *Bayes' theorem* (see subsection 30.2.3), we may write

$$P(\mathbf{a}|\mathbf{x}, H) = \frac{P(\mathbf{x}|\mathbf{a}, H)P(\mathbf{a}|H)}{P(\mathbf{x}|H)}, \tag{31.81}$$

which provides us with an expression for the probability distribution $P(\mathbf{a}|\mathbf{x}, H)$ of the parameters $\mathbf{a}$, given the (fixed) data $\mathbf{x}$ and our hypothesis H, in terms of other quantities that we may assign. The various terms in (31.81) have special formal names, as follows.

- The quantity $P(\mathbf{a}|H)$ on the RHS is the *prior* probability, which represents our state of knowledge of the parameter values (given the hypothesis H) *before* we have analysed the data.
- This probability is modified by the experimental data $\mathbf{x}$ through the *likelihood* $P(\mathbf{x}|\mathbf{a}, H)$.
- When appropriately normalised by the *evidence* $P(\mathbf{x}|H)$, this yields the *posterior* probability $P(\mathbf{a}|\mathbf{x}, H)$, which is the quantity of interest.
- The posterior encodes *all* our inferences about the values of the parameters $\mathbf{a}$. Strictly speaking, from a Bayesian viewpoint, this *entire function*, $P(\mathbf{a}|\mathbf{x}, H)$, is the 'answer' to a parameter estimation problem.

Given a particular hypothesis, the (normalising) evidence factor $P(\mathbf{x}|H)$ is unimportant, since it does not depend explicitly upon the parameter values $\mathbf{a}$. Thus, it is often omitted and one considers only the proportionality relation

$$P(\mathbf{a}|\mathbf{x}, H) \propto P(\mathbf{x}|\mathbf{a}, H)P(\mathbf{a}|H). \tag{31.82}$$

If necessary, the posterior distribution can be normalised empirically, by requiring that it integrates to unity, i.e. $\int P(\mathbf{a}|\mathbf{x}, H)\, d^m\mathbf{a} = 1$, where the integral extends over all values of the parameters $a_1, a_2, \ldots, a_m$.

The prior $P(\mathbf{a}|H)$ in (31.82) should reflect our entire knowledge concerning the values of the parameters $\mathbf{a}$, *before* the analysis of the current data $\mathbf{x}$. For example, there may be some physical reason to require some or all of the parameters to lie in a given range. If we are largely ignorant of the values of the parameters, we often indicate this by choosing a *uniform* (or very broad) prior,

$$P(\mathbf{a}|H) = \text{constant},$$

in which case the posterior distribution is simply proportional to the likelihood. In this case, we thus have

$$P(\mathbf{a}|\mathbf{x}, H) \propto L(\mathbf{x}; \mathbf{a}). \tag{31.83}$$

In other words, if we assume a uniform prior then we can identify the posterior distribution (up to a normalising factor) with $L(\mathbf{x}; \mathbf{a})$, considered as a function of the parameters $\mathbf{a}$.

Thus, a Bayesian statistician considers the ML estimates $\hat{\mathbf{a}}_{\mathrm{ML}}$ of the parameters to be the values that maximise the posterior $P(\mathbf{a}|\mathbf{x}, H)$ under the assumption of a uniform prior. More importantly, however, a Bayesian would *not* calculate the standard error or confidence interval on this estimate using the (classical) method employed in subsection 31.3.4. Instead, a far more straightforward approach is adopted. Let us assume, for the moment, that one is estimating just a single parameter a. Using (31.83), we may determine the values a_- and a_+ such that

$$\Pr(a < a_-|\mathbf{x}, H) = \int_{-\infty}^{a_-} L(\mathbf{x}; a)\, da = \alpha,$$
$$\Pr(a > a_+|\mathbf{x}, H) = \int_{a_+}^{\infty} L(\mathbf{x}; a)\, da = \beta.$$

where it is assumed that the likelihood has been normalised in such a way that $\int L(\mathbf{x}; a)\, da = 1$. Combining these equations gives

$$\Pr(a_- \leq a < a_+|\mathbf{x}, H) = \int_{a_-}^{a_+} L(\mathbf{x}; a)\, da = 1 - \alpha - \beta, \tag{31.84}$$

and $[a_-, a_+]$ is the *Bayesian confidence interval* on the value of a at the confidence level $1 - \alpha - \beta$. As in the case of classical confidence intervals, one often quotes the central confidence interval, for which $\alpha = \beta$. Another common choice (where possible) is to use the two values a_- and a_+ satisfying (31.84), for which $L(\mathbf{x}; a_-) = L(\mathbf{x}; a_+)$.

It should be understood that a frequentist would consider the Bayesian confidence interval as an *approximation* to the (classical) confidence interval discussed in subsection 31.3.4. Conversely, a Bayesian would consider the confidence interval defined in (31.84) to be the more meaningful. In fact, the difference between the Bayesian and classical confidence intervals is rather subtle. The classical confidence interval is defined in such a way that if one took a large number of samples each of size N and constructed the confidence interval in each case then the proportion of cases in which the true value of a would be contained within the interval is $1-\alpha-\beta$. For the Bayesian confidence interval, one does not rely on the frequentist concept of a large number of repeated samples. Instead, its meaning is that, given the single sample $\mathbf{x}$ (and our hypothesis H for the functional form of the population), the probability that a lies within the interval $[a_-, a_+]$ is $1-\alpha-\beta$.

By adopting the Bayesian viewpoint, the likelihood function $L(\mathbf{x}; a)$ may also be used to obtain an approximation $\hat{\sigma}_{\hat{a}}$ to the standard error in the ML estimator; the approximation is given by

$$\hat{\sigma}_{\hat{a}} = \left(- \left. \frac{\partial^2 \ln L}{\partial a^2} \right|_{a=\hat{a}} \right)^{-1/2}. \tag{31.85}$$

Clearly, if $L(\mathbf{x}; a)$ were a Gaussian centred on $a = \hat{a}$ then $\hat{\sigma}_{\hat{a}}$ would be its standard deviation. Indeed, in this case, the resulting 'one-sigma' limits would constitute a

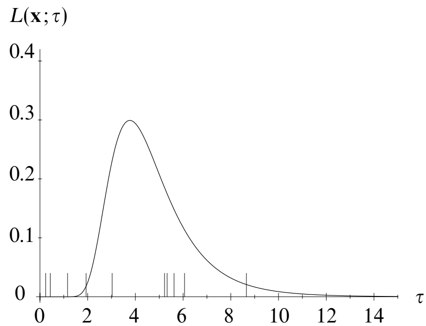


Figure 31.8 The likelihood function $L(\mathbf{x};\tau)$ (normalised to unit area) for the sample values given in the worked example in subsection 31.5.4 and indicated here by short vertical lines.

68.3% Bayesian central confidence interval. Even when $L(\mathbf{x};a)$ is not Gaussian, however, (31.85) is often used as a measure of the standard error.

► *For the sample data given in subsection 31.5.4, use the likelihood function to estimate the standard error $\hat{\sigma}_{\hat{\tau}}$ in the ML estimator $\hat{\tau}$ and obtain the Bayesian 68% central confidence interval on τ.*

We showed in (31.67) that the likelihood function in this case is given by

$$L(\mathbf{x};\tau) = \frac{1}{\tau^N}\exp\left[-\frac{1}{\tau}(x_1 + x_2 + \cdots + x_N)\right].$$

where x_i, $i = 1, 2, \ldots, N$, denotes the sample value and $N = 10$. This likelihood function is plotted in figure 31.8, after normalising (numerically) to unit area. The short vertical lines in the figure indicate the sample values. We see that the likelihood function peaks at the ML estimate $\hat{\tau} = 3.77$ that we found in subsection 31.5.4. Also, from (31.77), we have

$$\frac{\partial^2 \ln L}{\partial \tau^2} = \frac{N}{\tau}^2\left(1 - \frac{2}{\tau N}\sum_{i=1}^{N} x_i\right).$$

Remembering that $\hat{\tau} = \sum_i x_i/N$, our estimate of the standard error in $\hat{\tau}$ is

$$\hat{\sigma}_{\hat{\tau}} = \left(-\left.\frac{\partial^2 \ln L}{\partial \tau^2}\right|_{\tau=\hat{\tau}}\right)^{-1/2} = \frac{\hat{\tau}}{\sqrt{N}} = 1.19,$$

which is precisely the estimate of the standard error we obtained in subsection 31.5.4. It should be noted, however, that in general we would not expect the two estimates of standard error made by the different methods to be identical.

In order to calculate the Bayesian 68% central confidence interval, we must determine the values a_- and a_+ that satisfy (31.84) with $\alpha = \beta = 0.16$. In this case, the calculation can be performed analytically but is somewhat tedious. It is trivial, however, to determine a_- and a_+ numerically and we find the confidence interval to be $[3.16, 6.20]$. Thus we can quote our result with 68% central confidence limits as

$$\tau = 3.77^{+2.43}_{-0.61}.$$

By comparing this result with that given towards the end of subsection 31.5.4, we see that, as we might expect, the Bayesian and classical confidence intervals differ somewhat. ◄

The above discussion is generalised straightforwardly to the estimation of several parameters $a_1, a_2, \ldots, a_M$ simultaneously. The elements of the inverse of the covariance matrix of the ML estimators can be approximated by

$$(\mathsf{V}^{-1})_{ij} = -\left.\frac{\partial^2 \ln L}{\partial a_i \partial a_j}\right|_{\mathbf{a}=\hat{\mathbf{a}}}. \tag{31.86}$$

From (31.36), we see that (at least for unbiased estimators) the expectation value of (31.86) is equal to the element F_{ij} of the Fisher matrix.

The construction of a multi-dimensional *Bayesian confidence region* is also straightforward. For a given confidence level $1 - \alpha$ (say), it is most common to construct the confidence region as the M-dimensional region R in **a**-space, bounded by the 'surface' $L(\mathbf{x}; \mathbf{a}) =$ constant, for which

$$\int_R L(\mathbf{x}; \mathbf{a})\, d^M\mathbf{a} = 1 - \alpha,$$

where it is assumed that $L(\mathbf{x}; \mathbf{a})$ is normalised to unit volume. Moreover, we see from (31.83) that (assuming a uniform prior probability) we may obtain the *marginal* posterior distribution for any parameter a_i simply by integrating the likelihood function $L(\mathbf{x}; \mathbf{a})$ over the other parameters:

$$P(a_i|\mathbf{x}, H) = \int \cdots \int L(\mathbf{x}; \mathbf{a})\, da_1 \cdots da_{i-1} da_{i+1} \cdots da_M.$$

Here the integral extends over all possible values of the parameters, and again is it assumed that the likelihood function is normalised in such a way that $\int L(\mathbf{x}; \mathbf{a})\, d^M\mathbf{a} = 1$. This marginal distribution can then be used as above to determine Bayesian confidence intervals on each a_i separately.

►*Ten independent sample values x_i, $i = 1, 2, \ldots, 10$, are drawn at random from a Gaussian distribution with unknown mean μ and standard deviation σ. The sample values are as follows (to two decimal places):*

2.22 2.56 1.07 0.24 0.18 0.95 0.73 −0.79 2.09 1.81

Find the Bayesian 95% central confidence intervals on μ and σ separately.

The likelihood function in this case is

$$L(\mathbf{x}; \mu, \sigma) = (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2\right]. \tag{31.87}$$

Assuming uniform priors on μ and σ (over their natural ranges of $-\infty \to \infty$ and $0 \to \infty$ respectively), we may identify this likelihood function with the posterior probability, as in (31.83). Thus, the marginal posterior distribution on μ is given by

$$P(\mu|\mathbf{x}, H) \propto \int_0^\infty \frac{1}{\sigma^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2\right] d\sigma.$$

By substituting $\sigma = 1/u$ (so that $d\sigma = -du/u^2$) and integrating by parts either $(N-2)/2$ or $(N-3)/2$ times, we find

$$P(\mu|\mathbf{x}, H) \propto \left[N(\bar{x}-\mu)^2 + Ns^2\right]^{-(N-1)/2},$$

where we have used the fact that $\sum_i (x_i - \mu)^2 = N(\bar{x}-\mu)^2 + Ns^2$, $\bar{x}$ being the sample mean and s^2 the sample variance. We may now obtain the 95% central confidence interval by finding the values μ_- and μ_+ for which

$$\int_{-\infty}^{\mu_-} P(\mu|\mathbf{x}, H)\, d\mu = 0.025 \qquad \text{and} \qquad \int_{\mu_+}^{\infty} P(\mu|\mathbf{x}, H)\, d\mu = 0.025.$$

The normalisation of the posterior distribution and the values μ_- and μ_+ are easily obtained by numerical integration. Substituting in the appropriate values $N = 10$, $\bar{x} = 1.11$ and $s = 1.01$, we find the required confidence interval to be $[0.29, 1.97]$.

To obtain a confidence interval on σ, we must first obtain the corresponding marginal posterior distribution. From (31.87), again using the fact that $\sum_i (x_i-\mu)^2 = N(\bar{x}-\mu)^2+Ns^2$, this is given by

$$P(\sigma|\mathbf{x}, H) \propto \frac{1}{\sigma^N} \exp\left(-\frac{Ns^2}{2\sigma^2}\right) \int_{-\infty}^{\infty} \exp\left[-\frac{N(\bar{x}-\mu)^2}{2\sigma^2}\right] d\mu.$$

Noting that the integral of a one-dimensional Gaussian is proportional to σ, we conclude that

$$P(\sigma|\mathbf{x}, H) \propto \frac{1}{\sigma^{N-1}} \exp\left(-\frac{Ns^2}{2\sigma^2}\right).$$

The 95% central confidence interval on σ can then be found in an analogous manner to that on μ, by solving numerically the equations

$$\int_{0}^{\sigma_-} P(\sigma|\mathbf{x}, H)\, d\sigma = 0.025 \qquad \text{and} \qquad \int_{\sigma_+}^{\infty} P(\sigma|\mathbf{x}, H)\, d\sigma = 0.025.$$

We find the required interval to be $[0.76, 2.16]$. ◀

31.5.6 Behaviour of ML estimators for large N

As mentioned in subsection 31.3.6, in the large-sample limit $N \to \infty$, the sampling distribution of a set of (consistent) estimators $\hat{\mathbf{a}}$, whether ML or not, will tend, in general, to a multivariate Gaussian centred on the true values $\mathbf{a}$. This is a direct consequence of the central limit theorem. Similarly, in the limit $N \to \infty$ the likelihood function $L(\mathbf{x}; \mathbf{a})$ *also* tends towards a multivariate Gaussian but one centred on the ML estimate(s) $\hat{\mathbf{a}}$. Thus ML estimators are always *asymptotically consistent*. This limiting process was illustrated for the one-dimensional case by figure 31.5.

Thus, as N becomes large, the likelihood function tends to the form

$$L(\mathbf{x}; \mathbf{a}) = L_{\max} \exp\left[-\tfrac{1}{2} Q(\mathbf{a}, \hat{\mathbf{a}})\right],$$

where Q denotes the quadratic form

$$Q(\mathbf{a}, \hat{\mathbf{a}}) = (\mathbf{a} - \hat{\mathbf{a}})^{\mathrm{T}} \mathsf{V}^{-1} (\mathbf{a} - \hat{\mathbf{a}})$$

and the matrix V^{-1} is given by

$$\left(\mathsf{V}^{-1}\right)_{ij} = -\left.\frac{\partial^2 \ln L}{\partial a_i \partial a_j}\right|_{\mathbf{a}=\hat{\mathbf{a}}}.$$

Moreover, in the limit of large N, this matrix tends to the Fisher matrix given in (31.36), i.e. $\mathsf{V}^{-1} \to \mathsf{F}$. Hence ML estimators are *asymptotically minimum-variance.*

Comparison of the above results with those in subsection 31.3.6 shows that the large-sample limit of the likelihood function $L(\mathbf{x};\mathbf{a})$ has the same form as the large-sample limit of the joint estimator sampling distribution $P(\hat{\mathbf{a}}|\mathbf{a})$. The only difference is that $P(\hat{\mathbf{a}}|\mathbf{a})$ is centred in $\hat{\mathbf{a}}$-space on the true values $\hat{\mathbf{a}} = \mathbf{a}$ whereas $L(\mathbf{x};\mathbf{a})$ is centred in $\mathbf{a}$-space on the ML estimates $\mathbf{a} = \hat{\mathbf{a}}$. From figure 31.4 and its accompanying discussion, we therefore conclude that, in the large-sample limit, the Bayesian and classical confidence limits on the parameters *coincide.*

31.5.7 Extended maximum-likelihood method

It is sometimes the case that the number of data items N in our sample is itself a random variable. Such experiments are typically those in which data are collected for a certain period of time during which events occur at random in some way, as opposed to those in which a prearranged number of data items are collected. In particular, let us consider the case where the sample values $x_1, x_2, \ldots, x_N$ are drawn independently from some distribution $P(x|\mathbf{a})$ and the sample size N is a random variable described by a Poisson distribution with mean λ, i.e. $N \sim \text{Po}(\lambda)$. The likelihood function in this case is given by

$$L(\mathbf{x}, N; \lambda, \mathbf{a}) = \frac{\lambda^N}{N!} e^{-\lambda} \prod_{i=1}^{N} P(x_i|\mathbf{a}), \qquad (31.88)$$

and is often called the *extended likelihood function.* The function $L(\mathbf{x}; \lambda, \mathbf{a})$ can be used as before to estimate parameter values or obtain confidence intervals. Two distinct cases arise in the use of the extended likelihood function, depending on whether the Poisson parameter λ is a function of the parameters $\mathbf{a}$ or is an independent parameter.

Let us first consider the case in which λ is a function of the parameters $\mathbf{a}$. From (31.88), we can write the extended log-likelihood function as

$$\ln L = N \ln \lambda(\mathbf{a}) - \lambda(\mathbf{a}) + \sum_{i=1}^{N} \ln P(x_i|\mathbf{a}) = -\lambda(\mathbf{a}) + \sum_{i=1}^{N} \ln[\lambda(\mathbf{a}) P(x_i|\mathbf{a})].$$

where we have ignored terms not depending on $\mathbf{a}$. The ML estimates $\hat{\mathbf{a}}$ of the parameters can then be found in the usual way, and the ML estimate of the Poisson parameter is simply $\hat{\lambda} = \lambda(\hat{\mathbf{a}})$. The errors on our estimators $\hat{\mathbf{a}}$ will be, in general, smaller than those obtained in the usual likelihood approach, since our estimate includes information from the value of N as well as the sample values x_i.

The other possibility is that λ is an independent parameter and not a function of the parameters $\mathbf{a}$. In this case, the extended log-likelihood function is

$$\ln L = N \ln \lambda - \lambda + \sum_{i=1}^{N} \ln P(x_i|\mathbf{a}), \tag{31.89}$$

where we have omitted terms not depending on λ or $\mathbf{a}$. Differentiating with respect to λ and setting the result equal to zero, we find that the ML estimate of λ is simply

$$\hat{\lambda} = N.$$

By differentiating (31.89) with respect to the parameters a_i and setting the results equal to zero, we obtain the usual ML estimates $\hat{a}_i$ of their values. In this case, however, the errors in our estimates will be larger, in general, than those in the standard likelihood approach, since they must include the effect of statistical uncertainty in the parameter λ.

31.6 The method of least squares

The method of least squares is, in fact, just a special case of the method of maximum likelihood. Nevertheless, it is so widely used as a method of parameter estimation that it has acquired a special name of its own. At the outset, let us suppose that a data sample consists of a set of pairs (x_i, y_i), $i = 1, 2, \ldots, N$. For example, these data might correspond to the temperature y_i measured at various points x_i along some metal rod.

For the moment, we will suppose that the x_i are known exactly, whereas there exists a measurement error (or *noise*) n_i on each of the values y_i. Moreover, let us assume that the true value of y at any position x is given by some function $y = f(x; \mathbf{a})$ that depends on the M unknown parameters $\mathbf{a}$. Then

$$y_i = f(x_i; \mathbf{a}) + n_i.$$

Our aim is to estimate the values of the parameters $\mathbf{a}$ from the data sample.

Bearing in mind the central limit theorem, let us suppose that the n_i are drawn from a *Gaussian* distribution with no systematic bias and hence zero mean. In the most general case the measurement errors n_i might *not* be independent but be described by an N-dimensional multivariate Gaussian with non-trivial covariance matrix N, whose elements $N_{ij} = \mathrm{Cov}[n_i, n_j]$ we assume to be known. Under these assumptions it follows from (30.148), that the likelihood function is

$$L(\mathbf{x}, \mathbf{y}; \mathbf{a}) = \frac{1}{(2\pi)^{N/2}|\mathsf{N}|^{1/2}} \exp\left[-\tfrac{1}{2}\chi^2(\mathbf{a})\right],$$

where the quantity denoted by χ^2 is given by the quadratic form

$$\chi^2(\mathbf{a}) = \sum_{i,j=1}^{N} [y_i - f(x_i;\mathbf{a})](\mathsf{N}^{-1})_{ij}[y_j - f(x_j;\mathbf{a})] = (\mathsf{y}-\mathsf{f})^{\mathrm{T}}\mathsf{N}^{-1}(\mathsf{y}-\mathsf{f}). \qquad (31.90)$$

In the last equality, we have rewritten the expression in matrix notation by defining the column vector f with elements $f_i = f(x_i;\mathbf{a})$. We note that in the (common) special case in which the measurement errors n_i are *independent*, their covariance matrix takes the diagonal form $\mathsf{N} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$, where σ_i is the standard deviation of the measurement error n_i. In this case, the expression (31.90) for χ^2 reduces to

$$\chi^2(\mathbf{a}) = \sum_{i=1}^{N}\left[\frac{y_i - f(x_i;\mathbf{a})}{\sigma_i}\right]^2.$$

The least-squares (LS) estimators $\hat{\mathbf{a}}_{\mathrm{LS}}$ of the parameter values are defined as those that minimise the value of $\chi^2(\mathbf{a})$; they are usually determined by solving the M equations

$$\left.\frac{\partial \chi^2}{\partial a_i}\right|_{\mathbf{a}=\hat{\mathbf{a}}_{\mathrm{LS}}} = 0 \qquad \text{for } i = 1, 2, \ldots, M. \qquad (31.91)$$

Clearly, if the measurement errors n_i are indeed Gaussian distributed, as assumed above, then the LS and ML estimators of the parameters $\mathbf{a}$ coincide. Because of its relative simplicity, the method of least squares is often applied to cases in which the n_i are not Gaussian distributed. The resulting estimators $\hat{\mathbf{a}}_{\mathrm{LS}}$ are *not* the ML estimators, and the best that can be said in justification is that the method is an obviously sensible procedure for parameter estimation that has stood the test of time.

Finally, we note that the method of least squares is easily extended to the case in which each measurement y_i depends on several variables, which we denote by $\mathbf{x}_i$. For example, y_i might represent the temperature measured at the (three-dimensional) position $\mathbf{x}_i$ in a room. In this case, the data is modelled by a function $y = f(\mathbf{x}_i;\mathbf{a})$, and the remainder of the above discussion carries through unchanged.

31.6.1 Linear least squares

We have so far made no restriction on the form of the function $f(x;\mathbf{a})$. It so happens, however, that, for a model in which $f(x;\mathbf{a})$ is a *linear* function of the parameters $a_1, a_2, \ldots, a_M$, one can always obtain analytic expressions for the LS estimators $\hat{\mathbf{a}}_{\mathrm{LS}}$ and their variances. The general form of this kind of model is

$$f(x;\mathbf{a}) = \sum_{i=1}^{M} a_i h_i(x), \qquad (31.92)$$

where $\{h_1(x), h_2(x), \ldots, h_M(x)\}$ is some set of linearly independent fixed functions of x, often called the *basis functions*. Note that the functions $h_i(x)$ themselves may be highly non-linear functions of x. The 'linear' nature of the model (31.92) refers only to its dependence on the *parameters* a_i. Furthermore, in this case, it may be shown that the LS estimators $\hat{a}_i$ have zero bias and are minimum-variance, irrespective of the probability density function from which the measurement errors n_i are drawn.

In order to obtain analytic expressions for the LS estimators $\hat{\mathbf{a}}_{\rm LS}$, it is convenient to write (31.92) in the form

$$f(x;\mathbf{a}) = \sum_{j=1}^{M} R_{ij}a_j, \tag{31.93}$$

where $R_{ij} = h_j(x_i)$ is an element of the *response matrix* R of the experiment. The expression for χ^2 given in (31.90) can then be written, in matrix notation, as

$$\chi^2(\mathbf{a}) = (\mathsf{y} - \mathsf{Ra})^{\rm T}\mathsf{N}^{-1}(\mathsf{y} - \mathsf{Ra}). \tag{31.94}$$

The LS estimates of the parameters $\mathbf{a}$ are now found, as shown in (31.91), by differentiating (31.94) with respect to the a_i and setting the resulting expressions equal to zero. Denoting by $\nabla\chi^2$ the vector with elements $\partial\chi^2/\partial a_i$, we find

$$\nabla\chi^2 = -2\mathsf{R}^{\rm T}\mathsf{N}^{-1}(\mathsf{y} - \mathsf{Ra}). \tag{31.95}$$

This can be verified by writing out the expression (31.94) in component form and differentiating directly.

►*Verify result (31.95) by formulating the calculation in component form.*

To make the derivation less cumbersome, let us adopt the summation convention discussed in section 26.1, in which it is understood that any subscript that appears *exactly* twice in any term of an expression is to be summed over all the values that a subscript in that position can take. Thus, writing (31.94) in component form, we have

$$\chi^2(\mathbf{a}) = (y_i - R_{ik}a_k)(N^{-1})_{ij}(y_j - R_{jl}a_l).$$

Differentiating with respect to a_p gives

$$\begin{aligned}\frac{\partial\chi^2}{\partial a_p} &= -R_{ik}\delta_{kp}(N^{-1})_{ij}(y_j - R_{jl}a_l) + (y_i - R_{ik}a_k)(N^{-1})_{ij}(-R_{jl}\delta_{lp})\\ &= -R_{ip}(N^{-1})_{ij}(y_j - R_{jl}a_l) - (y_i - R_{ik}a_k)(N^{-1})_{ij}R_{jp},\end{aligned} \tag{31.96}$$

where δ_{ij} is the Kronecker delta symbol discussed in section 26.1. By swapping the indices i and j in the second term on the RHS of (31.96) and using the fact that the matrix N^{-1} is symmetric, we obtain

$$\begin{aligned}\frac{\partial\chi^2}{\partial a_p} &= -2R_{ip}(N^{-1})_{ij}(y_j - R_{jk}a_k)\\ &= -2(R^{\rm T})_{pi}(N^{-1})_{ij}(y_j - R_{jk}a_k).\end{aligned} \tag{31.97}$$

If we denote the vector with components $\partial\chi^2/\partial a_p$, $p = 1, 2, \ldots, M$, by $\nabla\chi^2$ and write the RHS of (31.97) in matrix notation, we recover the result (31.95). ◄

Setting the expression (31.95) equal to zero at $\mathbf{a} = \hat{\mathbf{a}}$, we find

$$-2\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{y} + 2\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R}\hat{\mathsf{a}} = 0.$$

Provided the matrix $\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R}$ is not singular, we may solve this equation for $\hat{\mathbf{a}}$ to obtain

$$\hat{\mathsf{a}} = (\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R})^{-1}\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{y} \equiv \mathsf{S}\mathsf{y}, \tag{31.98}$$

thus defining the $M \times N$ matrix S. It follows that the LS estimates $\hat{a}_i$, $i = 1, 2, \ldots, M$, are linear functions of the original measurements y_j, $j = 1, 2, \ldots, N$. Moreover, using the error propagation formula (30.141) derived in subsection 30.12.3, we find that the covariance matrix of the estimators $\hat{a}_i$ is given by

$$\mathsf{V} \equiv \mathrm{Cov}[\hat{a}_i, \hat{a}_j] = \mathsf{S}\mathsf{N}\mathsf{S}^{\mathrm{T}} = (\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R})^{-1}. \tag{31.99}$$

The two equations (31.98) and (31.99) contain the complete method of least squares. In particular, we note that, if one calculates the LS estimates using (31.98) then one has already obtained their covariance matrix (31.99).

►*Prove result (31.99).*

Using the definition of S given in (31.98), the covariance matrix (31.99) becomes

$$\begin{aligned}\mathsf{V} &= \mathsf{S}\mathsf{N}\mathsf{S}^{\mathrm{T}}\\ &= [(\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R})^{-1}\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}]\mathsf{N}[(\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R})^{-1}\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}]^{\mathrm{T}}.\end{aligned}$$

Using the result $(\mathsf{A}\mathsf{B}\cdots\mathsf{C})^{\mathrm{T}} = \mathsf{C}^{\mathrm{T}}\cdots\mathsf{B}^{\mathrm{T}}\mathsf{A}^{\mathrm{T}}$ for the transpose of a product of matrices and noting that, for any non-singular matrix, $(\mathsf{A}^{-1})^{\mathrm{T}} = (\mathsf{A}^{\mathrm{T}})^{-1}$ we find

$$\begin{aligned}\mathsf{V} &= (\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R})^{-1}\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{N}(\mathsf{N}^{\mathrm{T}})^{-1}\mathsf{R}[(\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R})^{\mathrm{T}}]^{-1}\\ &= (\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R})^{-1}\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R}(\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R})^{-1}\\ &= (\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R})^{-1},\end{aligned}$$

where we have also used the fact that N is symmetric and so $\mathsf{N}^{\mathrm{T}} = \mathsf{N}$. ◄

It is worth noting that one may also write the elements of the (inverse) covariance matrix as

$$(V^{-1})_{ij} = \frac{1}{2}\left(\frac{\partial^2\chi^2}{\partial a_i \partial a_j}\right)_{\mathbf{a}=\hat{\mathbf{a}}},$$

which is the same as the Fisher matrix (31.36) in cases where the measurement errors are Gaussian distributed (and so the log-likelihood is $\ln L = -\chi^2/2$). This proves, at least for this case, our earlier statement that the LS estimators are minimum-variance. In fact, since $f(x; \mathbf{a})$ is linear in the parameters $\mathbf{a}$, one can write χ^2 *exactly* as

$$\chi^2(\mathbf{a}) = \chi^2(\hat{\mathbf{a}}) + \frac{1}{2}\sum_{i,j=1}^{M}\left(\frac{\partial^2\chi^2}{\partial a_i \partial a_j}\right)_{\mathbf{a}=\hat{\mathbf{a}}}(a_i - \hat{a}_i)(a_j - \hat{a}_j),$$

which is quadratic in the parameters a_i. Hence the form of the likelihood function

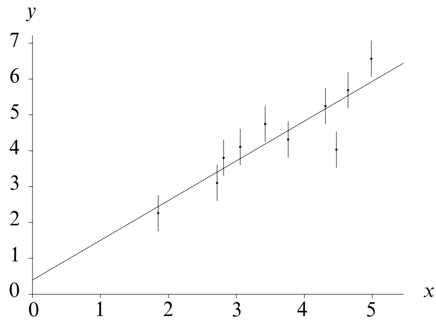


Figure 31.9 A set of data points with error bars indicating the uncertainty $\sigma = 0.5$ on the y-values. The straight line is $y = \hat{m}x + \hat{c}$, where $\hat{m}$ and $\hat{c}$ are the least-squares estimates of the slope and intercept.

$L \propto \exp(-\chi^2/2)$ is Gaussian. From the discussions of subsections 31.3.6 and 31.5.6, it follows that the 'surfaces' $\chi^2(\mathbf{a}) = c$, where c is a constant, bound ellipsoidal *confidence regions* for the parameters a_i. The relationship between the value of the constant c and the confidence level is given by (31.39).

►*An experiment produces the following data sample pairs* (x_i, y_i)*:*

x_i:	1.85	2.72	2.81	3.06	3.42	3.76	4.31	4.47	4.64	4.99
y_i:	2.26	3.10	3.80	4.11	4.74	4.31	5.24	4.03	5.69	6.57

where the x_i*-values are known exactly but each* y_i*-value is measured only to an accuracy of* $\sigma = 0.5$. *Assuming the underlying model for the data to be a straight line* $y = mx + c$, *find the LS estimates of the slope* m *and intercept* c *and quote the standard error on each estimate.*

The data are plotted in figure 31.9, together with error bars indicating the uncertainty in the y_i-values. Our model of the data is a straight line, and so we have

$$f(x; c, m) = c + mx.$$

In the language of (31.92), our basis functions are $h_1(x) = 1$ and $h_2(x) = x$ and our model parameters are $a_1 = c$ and $a_2 = m$. From (31.93) the elements of the response matrix are $R_{ij} = h_j(x_i)$, so that

$$\mathsf{R} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}, \tag{31.100}$$

where x_i are the data values and $N = 10$ in our case. Further, since the standard deviation on each measurement error is σ, we have $\mathsf{N} = \sigma^2 \mathsf{I}$, where I is the $N \times N$ identity matrix. Because of this simple form for N, the expression (31.98) for the LS estimates reduces to

$$\hat{\mathsf{a}} = \sigma^2(\mathsf{R}^{\mathrm{T}}\mathsf{R})^{-1}\frac{1}{\sigma^2}\mathsf{R}^{\mathrm{T}}\mathsf{y} = (\mathsf{R}^{\mathrm{T}}\mathsf{R})^{-1}\mathsf{R}^{\mathrm{T}}\mathsf{y}. \tag{31.101}$$

Note that we cannot expand the inverse in the last line, since R itself is not square and hence does not possess an inverse. Inserting the form for R in (31.100) into the expression (31.101), we find

$$\begin{pmatrix} \hat{c} \\ \hat{m} \end{pmatrix} = \begin{pmatrix} \sum_i 1 & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}$$
$$= \frac{1}{N(\overline{x^2} - \bar{x}^2)} \begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} N\bar{y} \\ N\overline{xy} \end{pmatrix}.$$

We thus obtain the LS estimates

$$\hat{m} = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - \bar{x}^2} \qquad \text{and} \qquad \hat{c} = \frac{\overline{x^2}\bar{y} - \bar{x}\,\overline{xy}}{\overline{x^2} - \bar{x}^2} = \bar{y} - \hat{m}\bar{x}, \tag{31.102}$$

where the last expression for $\hat{c}$ shows that the best-fit line passes through the 'centre of mass' $(\bar{x}, \bar{y})$ of the data sample. To find the standard errors on our results, we must calculate the covariance matrix of the estimators. This is given by (31.99), which in our case reduces to

$$\mathsf{V} = \sigma^2(\mathsf{R}^{\mathrm{T}}\mathsf{R})^{-1} = \frac{\sigma^2}{N(\overline{x^2} - \bar{x}^2)} \begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}. \tag{31.103}$$

The standard error on each estimator is simply the positive square root of the corresponding diagonal element, i.e. $\sigma_{\hat{c}} = \sqrt{V_{11}}$ and $\sigma_{\hat{m}} = \sqrt{V_{22}}$, and the covariance of the estimators $\hat{m}$ and $\hat{c}$ is given by $\mathrm{Cov}[\hat{c}, \hat{m}] = V_{12} = V_{21}$. Inserting the data sample averages and moments into (31.102) and (31.103), we find

$$c = \hat{c} \pm \sigma_{\hat{c}} = 0.40 \pm 0.62 \qquad \text{and} \qquad m = \hat{m} \pm \sigma_{\hat{m}} = 1.11 \pm 0.17.$$

The 'best-fit' straight line $y = \hat{m}x + \hat{c}$ is plotted in figure 31.9. For comparison, the true values used to create the data were $m = 1$ and $c = 1$. ◄

The extension of the method to fitting data to a higher-order polynomial, such as $f(x; \mathbf{a}) = a_1 + a_2 x + a_3 x^2$, is obvious. However, as the order of the polynomial increases the matrix inversions become rather complicated. Indeed, even when the matrices are inverted numerically, the inversion is prone to numerical instabilities. A better approach is to replace the basis functions $h_m(x) = x^m$, $m = 1, 2, \ldots, M$, with a set of polynomials that are 'orthogonal over the data', i.e. such that

$$\sum_{i=1}^{N} h_l(x_i) h_m(x_i) = 0 \qquad \text{for } l \neq m.$$

Such a set of polynomial basis functions can always be found by using the Gram–Schmidt orthogonalisation procedure presented in section 17.1. The details of this approach are beyond the scope of our discussion but we note that, in this case, the matrix $\mathsf{R}^{\mathrm{T}}\mathsf{R}$ is diagonal and may be inverted easily.

31.6.2 Non-linear least squares

If the function $f(x; \mathbf{a})$ is *not* linear in the parameters $\mathbf{a}$ then, in general, it is not possible to obtain an explicit expression for the LS estimates $\hat{\mathbf{a}}$. Instead, one must use an iterative (numerical) procedure, which we now outline. In practice, however, such problems are best solved using one of the many commercially available software packages.

One begins by making a first guess $\mathbf{a}^0$ for the values of the parameters. At this point in parameter space, the components of the gradient $\nabla\chi^2$ will not be equal to zero, in general (unless one makes a very lucky guess!). Thus, for at least some values of i, we have

$$\left.\frac{\partial\chi^2}{\partial a_i}\right|_{\mathbf{a}=\mathbf{a}^0} \neq 0.$$

Our aim is to find a small increment $\delta\mathbf{a}$ in the values of the parameters, such that

$$\left.\frac{\partial\chi^2}{\partial a_i}\right|_{\mathbf{a}=\mathbf{a}^0+\delta\mathbf{a}} = 0 \qquad \text{for all } i. \tag{31.104}$$

If our first guess $\mathbf{a}^0$ were sufficiently close to the true (local) minimum of χ^2, we could find the required increment $\delta\mathbf{a}$ by expanding the LHS of (31.104) as a Taylor series about $\mathbf{a} = \mathbf{a}^0$, keeping only the zeroth-order and first-order terms:

$$\left.\frac{\partial\chi^2}{\partial a_i}\right|_{\mathbf{a}=\mathbf{a}^0+\delta\mathbf{a}} \approx \left.\frac{\partial\chi^2}{\partial a_i}\right|_{\mathbf{a}=\mathbf{a}^0} + \sum_{j=1}^{M} \left.\frac{\partial^2\chi^2}{\partial a_i\partial a_j}\right|_{\mathbf{a}=\mathbf{a}^0} \delta a_j. \tag{31.105}$$

Setting this expression to zero, we find that the increments δa_j may be found by solving the set of M linear equations

$$\sum_{j=1}^{M} \left.\frac{\partial^2\chi^2}{\partial a_i\partial a_j}\right|_{\mathbf{a}=\mathbf{a}^0} \delta a_j = -\left.\frac{\partial\chi^2}{\partial a_i}\right|_{\mathbf{a}=\mathbf{a}^0}.$$

It most cases, however, our first guess $\mathbf{a}^0$ will not be sufficiently close to the true minimum for (31.105) to be an accurate approximation, and consequently (31.104) will not be satisfied. In this case, $\mathbf{a}^1 = \mathbf{a}^0 + \delta\mathbf{a}$ is (hopefully) an improved guess at the parameter values; the whole process is then repeated until convergence is achieved.

It is worth noting that, when one is estimating several parameters $\mathbf{a}$, the function $\chi^2(\mathbf{a})$ may be *very* complicated. In particular, it may possess numerous local extrema. The procedure outlined above will converge to the local extremum 'nearest' to the first guess $\mathbf{a}^0$. Since, in fact, we are interested only in the local minimum that has the absolute lowest value of $\chi^2(\mathbf{a})$, it is clear that a large part of solving the problem is to make a 'good' first guess.

31.7 Hypothesis testing

So far we have concentrated on using a data sample to obtain a number or a set of numbers. These numbers may be estimated values for the moments or central moments of the population from which the sample was drawn or, more generally, the values of some parameters $\mathbf{a}$ in an assumed model for the data. Sometimes, however, one wishes to use the data to give a 'yes' or 'no' answer to a particular question. For example, one might wish to know whether some assumed model does, in fact, provide a good fit to the data, or whether two parameters have the same value.

31.7.1 Simple and composite hypotheses

In order to use data to answer questions of this sort, the question must be posed precisely. This is done by first asserting that some *hypothesis* is true. The hypothesis under consideration is traditionally called the *null hypothesis* and is denoted by H_0. In particular, this usually specifies some form $P(\mathbf{x}|H_0)$ for the probability density function from which the data $\mathbf{x}$ are drawn. If the hypothesis determines the PDF uniquely, then it is said to be a *simple hypothesis*. If, however, the hypothesis determines the functional form of the PDF but not the values of certain parameters $\mathbf{a}$ on which it depends then it is called a *composite hypothesis*.

One decides whether to *accept* or *reject* the null hypothesis H_0 by performing some *statistical test*, as described below in subsection 31.7.2. In fact, formally one uses a statistical test to decide between the null hypothesis H_0 and the *alternative hypothesis* H_1. We define the latter to be the complement $\overline{H}_0$ of the null hypothesis *within some restricted hypothesis space known (or assumed) in advance*. Hence, rejection of H_0 implies acceptance of H_1, and vice versa.

As an example, let us consider the case in which a sample $\mathbf{x}$ is drawn from a Gaussian distribution with a known variance σ^2 but with an unknown mean μ. If one adopts the null hypothesis H_0 that $\mu = 0$, which we write as $H_0 : \mu = 0$, then the corresponding alternative hypothesis must be $H_1 : \mu \neq 0$. Note that, in this case, H_0 is a simple hypothesis whereas H_1 is a composite hypothesis. If, however, one adopted the null hypothesis $H_0 : \mu < 0$ then the alternative hypothesis would be $H_1 : \mu \geq 0$, so that both H_0 and H_1 would be composite hypotheses. Very occasionally both H_0 and H_1 will be simple hypotheses. In our illustration, this would occur, for example, if one knew in advance that the mean μ of the Gaussian distribution were equal to either zero or unity. In this case, if one adopted the null hypothesis $H_0 : \mu = 0$ then the alternative hypothesis would be $H_1 : \mu = 1$.

31.7.2 Statistical tests

In our discussion of hypothesis testing we will restrict our attention to cases in which the null hypothesis H_0 is *simple* (see above). We begin by constructing a *test statistic* $t(\mathbf{x})$ from the data sample. Although, in general, the test statistic need not be just a (scalar) number, and could be a multi-dimensional (vector) quantity, we will restrict our attention to the former case. Like any statistic, $t(\mathbf{x})$ will be a

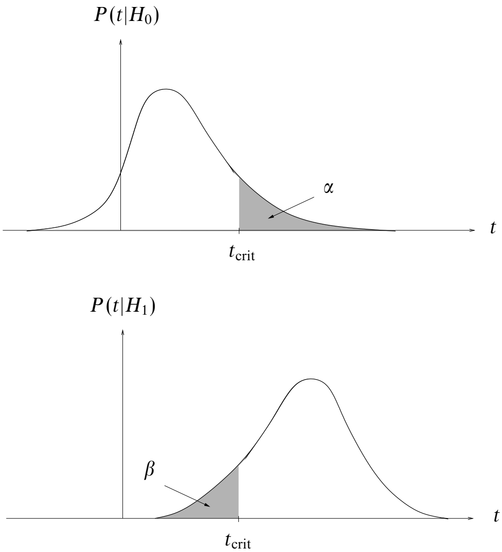


Figure 31.10 The sampling distributions $P(t|H_0)$ and $P(t|H_1)$ of a test statistic t. The shaded areas indicate the (one-tailed) regions for which $\Pr(t > t_{\text{crit}}|H_0) = \alpha$ and $\Pr(t < t_{\text{crit}}|H_1) = \beta$ respectively.

random variable. Moreover, given the simple null hypothesis H_0 concerning the PDF from which the sample was drawn, we may determine (in principle) the sampling distribution $P(t|H_0)$ of the test statistic. A typical example of such a sampling distribution is shown in figure 31.10. One defines for t a *rejection region* containing some fraction α of the total probability. For example, the (one-tailed) rejection region could consist of values of t greater than some value t_{crit}, for which

$$\Pr(t > t_{\text{crit}}|H_0) = \int_{t_{\text{crit}}}^{\infty} P(t|H_0)\,dt = \alpha; \qquad (31.106)$$

this is indicated by the shaded region in the upper half of figure 31.10. Equally, a (one-tailed) rejection region could consist of values of t less than some value t_{crit}. Alternatively, one could define a (two-tailed) rejection region by two values t_1 and t_2 such that $\Pr(t_1 < t < t_2|H_0) = \alpha$. In all cases, if the observed value of t lies in the rejection region then H_0 is *rejected* at *significance level* α; otherwise H_0 is *accepted* at this same level.

It is clear that there is a probability α of rejecting the null hypothesis H_0 even if it is true. This is called an *error of the first kind*. Conversely, an *error of the second kind* occurs when the hypothesis H_0 is accepted even though it is false (in which case H_1 is true). The probability β (say) that such an error will occur is, in general, difficult to calculate, since the alternative hypothesis H_1 is often composite. Nevertheless, in the case where H_1 is a simple hypothesis, it is straightforward (in principle) to calculate β. Denoting the corresponding sampling distribution of t by $P(t|H_1)$, the probability β is the integral of $P(t|H_1)$ over the *complement* of the rejection region, called the *acceptance region*. For example, in the case corresponding to (31.106) this probability is given by

$$\beta = \Pr(t < t_{\rm crit}|H_1) = \int_{-\infty}^{t_{\rm crit}} P(t|H_1)\,dt.$$

This is illustrated in figure 31.10. The quantity $1 - \beta$ is called the *power* of the statistical test to reject the wrong hypothesis.

31.7.3 The Neyman–Pearson test

In the case where H_0 and H_1 are both simple hypotheses, the *Neyman–Pearson lemma* (which we shall not prove) allows one to determine the 'best' rejection region and test statistic to use.

We consider first the choice of rejection region. Even in the general case, in which the test statistic $\mathbf{t}$ is a multi-dimensional (vector) quantity, the Neyman–Pearson lemma states that, for a given significance level α, the rejection region for H_0 giving the highest power for the test is the region of t-space for which

$$\frac{P(\mathbf{t}|H_0)}{P(\mathbf{t}|H_1)} > c, \tag{31.107}$$

where c is some constant determined by the required significance level.

In the case where the test statistic t is a simple scalar quantity, the Neyman–Pearson lemma is also useful in deciding which such statistic is the 'best' in the sense of having the maximum power for a given significance level α. From (31.107), we can see that the best statistic is given by the *likelihood ratio*

$$t(\mathbf{x}) = \frac{P(\mathbf{x}|H_0)}{P(\mathbf{x}|H_1)}. \tag{31.108}$$

and that the corresponding rejection region for H_0 is given by $t < t_{\rm crit}$. In fact, it is clear that any statistic $u = f(t)$ will be equally good, provided that $f(t)$ is a monotonically increasing function of t. The rejection region is then $u < f(t_{\rm crit})$. Alternatively, one may use any test statistic $v = g(t)$ where $g(t)$ is a monotonically decreasing function of t; in this case the rejection region becomes $v > g(t_{\rm crit})$. To construct such statistics, however, one must know $P(\mathbf{x}|H_0)$ and $P(\mathbf{x}|H_1)$ explicitly, and such cases are rare.

►*Ten independent sample values x_i, $i = 1, 2, \ldots, 10$, are drawn at random from a Gaussian distribution with standard deviation $\sigma = 1$. The mean μ of the distribution is known to equal either zero or unity. The sample values are as follows:*

2.22	2.56	1.07	0.24	0.18	0.95	0.73	−0.79	2.09	1.81

Test the null hypothesis $H_0 : \mu = 0$ at the 10% significance level.

The restricted nature of the hypothesis space means that our null and alternative hypotheses are $H_0 : \mu = 0$ and $H_1 : \mu = 1$ respectively. Since H_0 and H_1 are both simple hypotheses, the best test statistic is given by the likelihood ratio (31.108). Thus, denoting the means by μ_0 and μ_1, we have

$$\begin{aligned} t(\mathbf{x}) &= \frac{\exp\left[-\frac{1}{2}\sum_i (x_i - \mu_0)^2\right]}{\exp\left[-\frac{1}{2}\sum_i (x_i - \mu_1)^2\right]} = \frac{\exp\left[-\frac{1}{2}\sum_i (x_i^2 - 2\mu_0 x_i + \mu_0^2)\right]}{\exp\left[-\frac{1}{2}\sum_i (x_i^2 - 2\mu_1 x_i + \mu_1^2)\right]} \\ &= \exp\left[(\mu_0 - \mu_1)\textstyle\sum_i x_i - \frac{1}{2}N(\mu_0^2 - \mu_1^2)\right]. \end{aligned}$$

Inserting the values $\mu_0 = 0$ and $\mu_1 = 1$, yields $t = \exp(-N\bar{x} + \frac{1}{2}N)$, where $\bar{x}$ is the sample mean. Since $-\ln t$ is a monotonically decreasing function of t, however, we may equivalently use as our test statistic

$$v = -\frac{1}{N}\ln t + \tfrac{1}{2} = \bar{x},$$

where we have divided by the sample size N and added $\frac{1}{2}$ for convenience. Thus we may take the sample mean as our test statistic. From (31.13), we know that the sampling distribution of the sample mean under our null hypothesis H_0 is the Gaussian distribution $N(\mu_0, \sigma^2/N)$, where $\mu_0 = 0$, $\sigma^2 = 1$ and $N = 10$. Thus $\bar{x} \sim N(0, 0.1)$.

Since $\bar{x}$ is a monotonically decreasing function of t, our best rejection region for a given significance α is $\bar{x} > \bar{x}_{\rm crit}$, where $\bar{x}_{\rm crit}$ depends on α. Thus, in our case, $\bar{x}_{\rm crit}$ is given by

$$\alpha = 1 - \Phi\left(\frac{\bar{x}_{\rm crit} - \mu_0}{\sigma}\right) = 1 - \Phi(10\bar{x}_{\rm crit}),$$

where $\Phi(z)$ is the cumulative distribution function for the standard Gaussian. For a 10% significance level we have $\alpha = 0.1$ and, from table 30.3 in subsection 30.9.1, we find $\bar{x}_{\rm crit} = 0.128$. Thus the rejection region on $\bar{x}$ is

$$\bar{x} > 0.128.$$

From the sample, we deduce that $\bar{x} = 1.11$, and so we can clearly reject the null hypothesis $H_0 : \mu = 0$ at the 10% significance level. It can, in fact, be rejected at a much higher significance level. As revealed on p. 1239, the data was generated using $\mu = 1$. ◄

31.7.4 The generalised likelihood-ratio test

If the null hypothesis H_0 or the alternative hypothesis H_1 is composite (or both are composite) then the corresponding distributions $P(\mathbf{x}|H_0)$ and $P(\mathbf{x}|H_1)$ are not uniquely determined, in general, and so we cannot use the Neyman–Pearson lemma to obtain the 'best' test statistic t. Nevertheless, in many cases, there still exists a general procedure for constructing a test statistic t which has useful properties and which reduces to the Neyman–Pearson statistic (31.108) in the special case where H_0 and H_1 are both simple hypotheses.

Consider the quite general, and commonly occurring, case in which the data sample $\mathbf{x}$ is drawn from a population $P(\mathbf{x}|\mathbf{a})$ with a known (or assumed) functional form but depends on the unknown values of some parameters $a_1, a_2, \ldots, a_M$. Moreover, suppose we wish to test the null hypothesis H_0 that the parameter values $\mathbf{a}$ lie in some subspace $\mathcal{S}$ of the full parameter space $\mathcal{A}$. In other words, on the basis of the sample $\mathbf{x}$ it is desired to test the null hypothesis $H_0 : (a_1, a_2, \ldots, a_M$ lies in $\mathcal{S})$ against the alternative hypothesis $H_1 : (a_1, a_2, \ldots, a_M$ lies in $\overline{\mathcal{S}})$, where $\overline{\mathcal{S}}$ is $\mathcal{A} - \mathcal{S}$.

Since the functional form of the population is known, we may write down the likelihood function $L(\mathbf{x}; \mathbf{a})$ for the sample. Ordinarily, the likelihood will have a maximum as the parameters $\mathbf{a}$ are varied over the entire parameter space $\mathcal{A}$. This is the usual maximum-likelihood estimate of the parameter values, which we denote by $\hat{\mathbf{a}}$. If, however, the parameter values are allowed to vary only over the subspace $\mathcal{S}$ then the likelihood function will be maximised at the point $\hat{\mathbf{a}}_{\mathcal{S}}$, which may or may not coincide with the global maximum $\hat{\mathbf{a}}$. Now, let us take as our test statistic the *generalised likelihood ratio*

$$t(\mathbf{x}) = \frac{L(\mathbf{x}; \hat{\mathbf{a}}_{\mathcal{S}})}{L(\mathbf{x}; \hat{\mathbf{a}})}, \tag{31.109}$$

where $L(\mathbf{x}; \hat{\mathbf{a}}_{\mathcal{S}})$ is the maximum value of the likelihood function in the subspace $\mathcal{S}$ and $L(\mathbf{x}; \hat{\mathbf{a}})$ is its maximum value in the entire parameter space $\mathcal{A}$. It is clear that t is a function of the sample values only and must lie between 0 and 1.

We will concentrate on the special case where H_0 is the simple hypothesis $H_0 : \mathbf{a} = \mathbf{a}_0$. The subspace $\mathcal{S}$ then consists of only the single point $\mathbf{a}_0$. Thus (31.109) becomes

$$t(\mathbf{x}) = \frac{L(\mathbf{x}; \mathbf{a}_0)}{L(\mathbf{x}; \hat{\mathbf{a}})}, \tag{31.110}$$

and the sampling distribution $P(t|H_0)$ can be determined (in principle). As in the previous subsection, the best rejection region for a given significance α is simply $t < t_{\rm crit}$, where the value $t_{\rm crit}$ depends on α. Moreover, as before, an equivalent procedure is to use as a test statistic $u = f(t)$, where $f(t)$ is any monotonically increasing function of t; the corresponding rejection region is then $u < f(t_{\rm crit})$. Similarly, one may use a test statistic $v = g(t)$, where $g(t)$ is any monotonically decreasing function of t; the rejection region then becomes $v > g(t_{\rm crit})$. Finally, we note that if H_1 is also a simple hypothesis $H_1 : \mathbf{a} = \mathbf{a}_1$, then (31.110) reduces to the Neyman–Pearson test statistic (31.108).

►*Ten independent sample values x_i, $i = 1, 2, \ldots, 10$, are drawn at random from a Gaussian distribution with standard deviation $\sigma = 1$. The sample values are as follows:*

2.22	2.56	1.07	0.24	0.18	0.95	0.73	−0.79	2.09	1.81

Test the null hypothesis $H_0 : \mu = 0$ at the 10% significance level.

We must test the (simple) null hypothesis $H_0 : \mu = 0$ against the (composite) alternative hypothesis $H_1 : \mu \neq 0$. Thus, the subspace $\mathcal{S}$ is the single point $\mu = 0$, whereas $\mathcal{A}$ is the entire μ-axis. The likelihood function is

$$L(\mathbf{x}; \mu) = \frac{1}{(2\pi)^{N/2}} \exp\left[-\tfrac{1}{2}\textstyle\sum_i (x_i - \mu)^2\right],$$

which has its global maximum at $\mu = \bar{x}$. The test statistic t is then given by

$$t(\mathbf{x}) = \frac{L(\mathbf{x}; 0)}{L(\mathbf{x}; \bar{x})} = \frac{\exp\left[-\frac{1}{2}\sum_i x_i^2\right]}{\exp\left[-\frac{1}{2}\sum_i (x_i - \bar{x})^2\right]} = \exp\left(-\tfrac{1}{2}N\bar{x}^2\right).$$

It is in fact more convenient to consider the test statistic

$$v = -2\ln t = N\bar{x}^2.$$

Since $-2\ln t$ is a monotonically decreasing function of t, the rejection region now becomes $v > v_{\rm crit}$, where

$$\int_{v_{\rm crit}}^{\infty} P(v|H_0)\, dv = \alpha, \qquad (31.111)$$

α being the significance level of the test. Thus it only remains to determine the sampling distribution $P(v|H_0)$. Under the null hypothesis H_0, we expect $\bar{x}$ to be Gaussian distributed, with mean zero and variance $1/N$. Thus, from subsection 30.9.4, v will follow a *chi-squared* distribution of order 1. Substituting the appropriate form for $P(v|H_0)$ in (31.111) and setting $\alpha = 0.1$, we find by numerical integration (or from table 31.2) that $v_{\rm crit} = N\bar{x}^2_{\rm crit} = 2.71$. Since $N = 10$, the rejection region on $\bar{x}$ at the 10% significance level is thus

$$\bar{x} < -0.52 \quad \text{and} \quad \bar{x} > 0.52.$$

As noted before, for this sample $\bar{x} = 1.11$, and so we may reject the null hypothesis $H_0 : \mu = 0$ at the 10% significance level. ◄

The above example illustrates the general situation that if the maximum-likelihood estimates $\hat{\mathbf{a}}$ of the parameters fall in or near the subspace $\mathcal{S}$ then the sample will be considered consistent with H_0 and the value of t will be near unity. If $\hat{\mathbf{a}}$ is distant from $\mathcal{S}$ then the sample will not be in accord with H_0 and ordinarily t will have a small (positive) value.

It is clear that in order to prescribe the rejection region for t, or for a related statistic u or v, it is necessary to know the sampling distribution $P(t|H_0)$. If H_0 is simple then one can in principle determine $P(t|H_0)$, although this may prove difficult in practice. Moreover, if H_0 is composite, then it may not be possible to obtain $P(t|H_0)$, even in principle. Nevertheless, a useful approximate form for $P(t|H_0)$ exists in the large-sample limit. Consider the null hypothesis

$$H_0 : (a_1 = a_1^0, a_2 = a_2^0, \ldots, a_R = a_R^0), \quad \text{where } R \leq M$$

and the a_i^0 are fixed numbers. (In fact, we may fix the values of any subset containing R of the M parameters.) If H_0 is true then it follows from our discussion in subsection 31.5.6 (although we shall not prove it) that, when the sample size N is large, the quantity $-2\ln t$ follows approximately a *chi-squared* distribution of order R.

31.7.5 Student's t-test

Student's t-test is just a special case of the generalised likelihood ratio test applied to a sample $x_1, x_2, \ldots, x_N$ drawn independently from a Gaussian distribution for which *both* the mean μ and variance σ^2 are unknown, and for which one wishes to distinguish between the hypotheses

$$H_0 : \mu = \mu_0, \quad 0 < \sigma^2 < \infty, \qquad \text{and} \qquad H_1 : \mu \neq \mu_0, \quad 0 < \sigma^2 < \infty,$$

where μ_0 is a given number. Here, the parameter space $\mathcal{A}$ is the half-plane $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$, whereas the subspace $\mathcal{S}$ characterised by the null hypothesis H_0 is the line $\mu = \mu_0$, $0 < \sigma^2 < \infty$.

The likelihood function for this situation is given by

$$L(\mathbf{x}; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{\sum_i (x_i - \mu)^2}{2\sigma^2}\right].$$

On the one hand, as shown in subsection 31.5.1, the values of μ and σ^2 that maximise L in $\mathcal{A}$ are $\mu = \bar{x}$ and $\sigma^2 = s^2$, where $\bar{x}$ is the sample mean and s^2 is the sample variance. On the other hand, to maximise L in the subspace $\mathcal{S}$ we set $\mu = \mu_0$, and the only remaining parameter is σ^2; the value of σ^2 that maximises L is then easily found to be

$$\widehat{\sigma^2} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_0)^2.$$

To retain, in due course, the standard notation for Student's t-test, in this section we will denote the generalised likelihood ratio by λ (rather than t); it is thus given by

$$\begin{aligned}
\lambda(\mathbf{x}) &= \frac{L(\mathbf{x}; \mu_0, \widehat{\sigma^2})}{L(\mathbf{x}; \bar{x}, s^2)} \\
&= \frac{[(2\pi/N)\sum_i (x_i - \mu_0)^2]^{-N/2} \exp(-N/2)}{[(2\pi/N)\sum_i (x_i - \bar{x})^2]^{-N/2} \exp(-N/2)} = \left[\frac{\sum_i (x_i - \bar{x})^2}{\sum_i (x_i - \mu_0)^2}\right]^{N/2}. \qquad (31.112)
\end{aligned}$$

Normally, our next step would be to find the sampling distribution of λ under the assumption that H_0 were true. It is more conventional, however, to work in terms of a related test statistic t, which was first devised by William Gossett, who wrote under the pen name of 'Student'.

The sum of squares in the denominator of (31.112) may be put into the form

$$\textstyle\sum_i(x_i-\mu_0)^2 = N(\bar{x}-\mu_0)^2 + \sum_i(x_i-\bar{x})^2.$$

Thus, on dividing the numerator and denominator in (31.112) by $\sum_i(x_i-\bar{x})^2$ and rearranging, the generalised likelihood ratio λ can be written

$$\lambda = \left(1+\frac{t^2}{N-1}\right)^{-N/2},$$

where we have defined the new variable

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{N-1}}. \tag{31.113}$$

Since t^2 is a monotonically decreasing function of λ, the corresponding rejection region is $t^2 > c$, where c is a positive constant depending on the required significance level α. It is conventional, however, to use t itself as our test statistic, in which case our rejection region becomes two-tailed and is given by

$$t < -t_{\rm crit} \qquad \text{and} \qquad t > t_{\rm crit}, \tag{31.114}$$

where $t_{\rm crit}$ is the positive square root of the constant c.

The definition (31.113) and the rejection region (31.114) form the basis of Student's t-test. It only remains to determine the sampling distribution $P(t|H_0)$. At the outset, it is worth noting that if we write the expression (31.113) for t in terms of the standard estimator $\hat{\sigma} = \sqrt{Ns^2/(N-1)}$ of the standard deviation then we obtain

$$t = \frac{\bar{x}-\mu_0}{\hat{\sigma}/\sqrt{N}}. \tag{31.115}$$

If, in fact, we knew the true value of σ and used it in this expression for t then it is clear from our discussion in section 31.3 that t would follow a Gaussian distribution with mean 0 and variance 1, i.e. $t \sim N(0,1)$. When σ is not known, however, we have to use our estimate $\hat{\sigma}$ in (31.115), with the result that t is no longer distributed as the standard Gaussian. As one might expect from the central limit theorem, however, the distribution of t does tend towards the standard Gaussian for large values of N.

As noted earlier, the exact distribution of t, valid for any value of N, was first discovered by William Gossett. From (31.35), if the hypothesis H_0 is true then the joint sampling distribution of $\bar{x}$ and s is given by

$$P(\bar{x},s|H_0) = Cs^{N-2}\exp\left(-\frac{Ns^2}{2\sigma^2}\right)\exp\left[-\frac{N(\bar{x}-\mu_0)^2}{2\sigma^2}\right], \tag{31.116}$$

where C is a normalisation constant. We can use this result to obtain the joint sampling distribution of s and t by demanding that

$$P(\bar{x},s|H_0)\,d\bar{x}\,ds = P(t,s|H_0)\,dt\,ds.$$

Using (31.113) to substitute for $\bar{x} - \mu_0$ in (31.116), and noting that $d\bar{x} = (s/\sqrt{N-1})\,dt$, we find

$$P(\bar{x}, s|H_0)\,d\bar{x}\,ds = As^{N-1}\exp\left[-\frac{Ns^2}{2\sigma^2}\left(1+\frac{t^2}{N-1}\right)\right]\,dt\,ds,$$

where A is another normalisation constant. In order to obtain the sampling distribution of t alone, we must integrate $P(t, s|H_0)$ with respect to s over its allowed range, from 0 to ∞. Thus, the required distribution of t alone is given by

$$P(t|H_0) = \int_0^\infty P(t, s|H_0)\,ds = A\int_0^\infty s^{N-1}\exp\left[-\frac{Ns^2}{2\sigma^2}\left(1+\frac{t^2}{N-1}\right)\right]\,ds. \tag{31.117}$$

To carry out this integration, we set $y = s\{1 + [t^2/(N-1)]\}^{1/2}$, which on substitution into (31.117) yields

$$P(t|H_0) = A\left(1+\frac{t^2}{N-1}\right)^{-N/2}\int_0^\infty y^{N-1}\exp\left(-\frac{Ny^2}{2\sigma^2}\right)\,dy.$$

Since the integral over y does not depend on t, it is simply a constant. We thus find that that the sampling distribution of the variable t is

$$P(t|H_0) = \frac{1}{\sqrt{(N-1)\pi}}\frac{\Gamma\left(\frac{1}{2}N\right)}{\Gamma\left(\frac{1}{2}(N-1)\right)}\left(1+\frac{t^2}{N-1}\right)^{-N/2}, \tag{31.118}$$

where we have used the condition $\int_{-\infty}^{\infty} P(t|H_0)\,dt = 1$ to determine the normalisation constant (see exercise 31.18).

The distribution (31.118) is called *Student's t-distribution with $N-1$ degrees of freedom*. A plot of Student's t-distribution is shown in figure 31.11 for various values of N. For comparison, we also plot the standard Gaussian distribution, to which the t-distribution tends for large N. As is clear from the figure, the t-distribution is symmetric about $t = 0$. In table 31.3 we list some critical points of the cumulative probability function $C_n(t)$ of the t-distribution, which is defined by

$$C_n(t) = \int_{-\infty}^{t} P(t'|H_0)\,dt',$$

where $n = N-1$ is the number of degrees of freedom. Clearly, $C_n(t)$ is analogous to the cumulative probability function $\Phi(z)$ of the Gaussian distribution, discussed in subsection 30.9.1. For comparison purposes, we also list the critical points of $\Phi(z)$, which corresponds to the t-distribution for $N = \infty$.

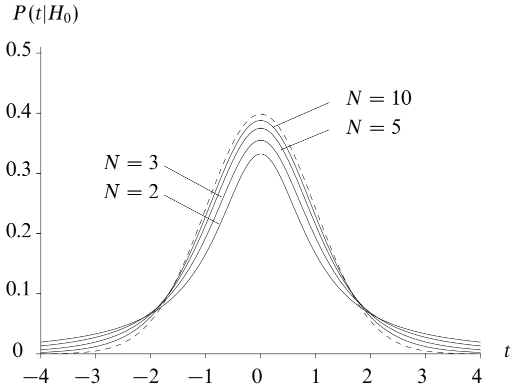


Figure 31.11 Student's t-distribution for various values of N. The broken curve shows the standard Gaussian distribution for comparison.

►*Ten independent sample values x_i, $i = 1, 2, \ldots, 10$, are drawn at random from a Gaussian distribution with unknown mean μ and unknown standard deviation σ. The sample values are as follows:*

2.22	2.56	1.07	0.24	0.18	0.95	0.73	−0.79	2.09	1.81

Test the null hypothesis $H_0 : \mu = 0$ at the 10% significance level.

For our null hypothesis, $\mu_0 = 0$. Since for this sample $\bar{x} = 1.11$, $s = 1.01$ and $N = 10$, it follows from (31.113) that

$$t = \frac{\bar{x}}{s/\sqrt{N-1}} = 3.33.$$

The rejection region for t is given by (31.114) where $t_{\rm crit}$ is such that

$$C_{N-1}(t_{\rm crit}) = 1 - \alpha/2,$$

and α is the required significance of the test. In our case $\alpha = 0.1$ and $N = 10$, and from table 31.3 we find $t_{\rm crit} = 1.83$. Thus our rejection region for H_0 at the 10% significance level is

$$t < -1.83 \qquad \text{and} \qquad t > 1.83.$$

For our sample $t = 3.30$ and so we can clearly reject the null hypothesis $H_0 : \mu = 0$ at this level. ◄

It is worth noting the connection between the t-test and the classical confidence interval on the mean μ. The central confidence interval on μ at the confidence level $1 - \alpha$ is the set of values for which

$$-t_{\rm crit} < \frac{\bar{x} - \mu}{s/\sqrt{N-1}} < t_{\rm crit},$$

$C_n(t)$	0.5	0.6	0.7	0.8	0.9	0.950	0.975	0.990	0.995	0.999
$n = 1$	0.00	0.33	0.73	1.38	3.08	6.31	12.7	31.8	63.7	318.3
2	0.00	0.29	0.62	1.06	1.89	2.92	4.30	6.97	9.93	22.3
3	0.00	0.28	0.58	0.98	1.64	2.35	3.18	4.54	5.84	10.2
4	0.00	0.27	0.57	0.94	1.53	2.13	2.78	3.75	4.60	7.17
5	0.00	0.27	0.56	0.92	1.48	2.02	2.57	3.37	4.03	5.89
6	0.00	0.27	0.55	0.91	1.44	1.94	2.45	3.14	3.71	5.21
7	0.00	0.26	0.55	0.90	1.42	1.90	2.37	3.00	3.50	4.79
8	0.00	0.26	0.55	0.89	1.40	1.86	2.31	2.90	3.36	4.50
9	0.00	0.26	0.54	0.88	1.38	1.83	2.26	2.82	3.25	4.30
10	0.00	0.26	0.54	0.88	1.37	1.81	2.23	2.76	3.17	4.14
11	0.00	0.26	0.54	0.88	1.36	1.80	2.20	2.72	3.11	4.03
12	0.00	0.26	0.54	0.87	1.36	1.78	2.18	2.68	3.06	3.93
13	0.00	0.26	0.54	0.87	1.35	1.77	2.16	2.65	3.01	3.85
14	0.00	0.26	0.54	0.87	1.35	1.76	2.15	2.62	2.98	3.79
15	0.00	0.26	0.54	0.87	1.34	1.75	2.13	2.60	2.95	3.73
16	0.00	0.26	0.54	0.87	1.34	1.75	2.12	2.58	2.92	3.69
17	0.00	0.26	0.53	0.86	1.33	1.74	2.11	2.57	2.90	3.65
18	0.00	0.26	0.53	0.86	1.33	1.73	2.10	2.55	2.88	3.61
19	0.00	0.26	0.53	0.86	1.33	1.73	2.09	2.54	2.86	3.58
20	0.00	0.26	0.53	0.86	1.33	1.73	2.09	2.53	2.85	3.55
25	0.00	0.26	0.53	0.86	1.32	1.71	2.06	2.49	2.79	3.46
30	0.00	0.26	0.53	0.85	1.31	1.70	2.04	2.46	2.75	3.39
40	0.00	0.26	0.53	0.85	1.30	1.68	2.02	2.42	2.70	3.31
50	0.00	0.26	0.53	0.85	1.30	1.68	2.01	2.40	2.68	3.26
100	0.00	0.25	0.53	0.85	1.29	1.66	1.98	2.37	2.63	3.17
200	0.00	0.25	0.53	0.84	1.29	1.65	1.97	2.35	2.60	3.13
∞	0.00	0.25	0.52	0.84	1.28	1.65	1.96	2.33	2.58	3.09

Table 31.3 The confidence limits t of the cumulative probability function $C_n(t)$ for Student's t-distribution with n degrees of freedom. For example, $C_5(0.92) = 0.8$. The row $n = \infty$ is also the corresponding result for the standard Gaussian distribution.

where t_{crit} satisfies $C_{N-1}(t_{\text{crit}}) = \alpha/2$. Thus the required confidence interval is

$$\bar{x} - \frac{t_{\text{crit}} s}{\sqrt{N-1}} < \mu < \bar{x} + \frac{t_{\text{crit}} s}{\sqrt{N-1}}.$$

Hence, in the above example, the 90% classical central confidence interval on μ is

$$0.49 < \mu < 1.73.$$

The t-distribution may also be used to compare different samples from Gaussian distributions. In particular, let us consider the case where we have two independent samples of sizes N_1 and N_2, drawn respectively from Gaussian distributions with a common variance σ^2 but with possibly different means μ_1 and μ_2. On the basis of the samples, one wishes to distinguish between the hypotheses

$$H_0 : \mu_1 = \mu_2, \quad 0 < \sigma^2 < \infty \qquad \text{and} \qquad H_1 : \mu_1 \neq \mu_2, \quad 0 < \sigma^2 < \infty.$$

In other words, we wish to test the null hypothesis that the samples are drawn from populations having the same mean. Suppose that the measured sample means and standard deviations are $\bar{x}_1$, $\bar{x}_2$ and s_1, s_2 respectively. In an analogous way to that presented above, one may show that the generalised likelihood ratio can be written as

$$\lambda = \left(1 + \frac{t^2}{N_1 + N_2 - 2}\right)^{-(N_1+N_2)/2}.$$

In this case, the variable t is given by

$$t = \frac{\bar{w} - \omega}{\hat{\sigma}} \left(\frac{N_1 N_2}{N_1 + N_2}\right)^{1/2}, \qquad (31.119)$$

where $\bar{w} = \bar{x}_1 - \bar{x}_2$, $\omega = \mu_1 - \mu_2$ and

$$\hat{\sigma} = \left[\frac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}\right]^{1/2}.$$

It is straightforward (albeit with complicated algebra) to show that the variable t in (31.119) follows Student's t-distribution with $N_1 + N_2 - 2$ degrees of freedom, and so we may use an appropriate form of Student's t-test to investigate the null hypothesis $H_0 : \mu_1 = \mu_2$ (or equivalently $H_0 : \omega = 0$). As above, the t-test can be used to place a confidence interval on $\omega = \mu_1 - \mu_2$.

►*Suppose that two classes of students take the same mathematics examination and the following percentage marks are obtained:*

Class 1:	66	62	34	55	77	80	55	60	69	47	50
Class 2:	64	90	76	56	81	72	70				

Assuming that the two sets of examinations marks are drawn from Gaussian distributions with a common variance, test the hypothesis $H_0 : \mu_1 = \mu_2$ at the 5% significance level. Use your result to obtain the 95% classical central confidence interval on $\omega = \mu_1 - \mu_2$.

We begin by calculating the mean and standard deviation of each sample. The number of values in each sample is $N_1 = 11$ and $N_2 = 7$ respectively, and we find

$$\bar{x}_1 = 59.5, \; s_1 = 12.8 \quad \text{and} \quad \bar{x}_2 = 72.7, \; s_2 = 10.3,$$

leading to $\bar{w} = \bar{x}_1 - \bar{x}_2 = -13.2$ and $\hat{\sigma} = 12.6$. Setting $\omega = 0$ in (31.119), we thus find $t = -2.17$.

The rejection region for H_0 is given by (31.114), where $t_{\rm crit}$ satisfies

$$C_{N_1+N_2-2}(t_{\rm crit}) = 1 - \alpha/2, \qquad (31.120)$$

where α is the required significance level of the test. In our case we set $\alpha = 0.05$, and from table 31.3 with $n = 16$ we find that $t_{\rm crit} = 2.12$. The rejection region is therefore

$$t < -2.12 \qquad \text{and} \qquad t > 2.12.$$

Since $t = -2.17$ for our samples, we can reject the null hypothesis $H_0 : \mu_1 = \mu_2$, although only by a small margin. (Indeed, it is easily shown that one cannot reject H_0 at the 2% significance level). The 95% central confidence interval on $\omega = \mu_1 - \mu_2$ is given by

$$\bar{w} - \hat{\sigma} t_{\rm crit} \left(\frac{N_1 + N_2}{N_1 N_2} \right)^{1/2} < \quad \omega \quad < \bar{w} + \hat{\sigma} t_{\rm crit} \left(\frac{N_1 + N_2}{N_1 N_2} \right)^{1/2},$$

where $t_{\rm crit}$ is given by (31.120). Thus, we find

$$-26.1 < \omega < -0.28,$$

which, as expected, does not (quite) contain $\omega = 0$. ◄

In order to apply Student's t-test in the above example, we had to make the assumption that the samples were drawn from Gaussian distributions possessing a common variance, which is clearly unjustified *a priori*. We can, however, perform another test on the data to investigate whether the additional hypothesis $\sigma_1^2 = \sigma_2^2$ is reasonable; this test is discussed in the next subsection. If this additional test shows that the hypothesis $\sigma_1^2 = \sigma_2^2$ may be accepted (at some suitable significance level), then we may indeed use the analysis in the above example to infer that the null hypothesis $H_0 : \mu_1 = \mu_2$ may be rejected at the 5% significance level. If, however, we find that the additional hypothesis $\sigma_1^2 = \sigma_2^2$ must be rejected, then we can only infer from the above example that the hypothesis that the two samples were drawn from the same Gaussian distribution may be rejected at the 5% significance level.

Throughout the above discussion, we have assumed that samples are drawn from a Gaussian distribution. Although this is true for many random variables, in practice it is usually impossible to know *a priori* whether this is case. It can be shown, however, that Student's t-test remains reasonably accurate even if the sampled distribution(s) differ considerably from a Gaussian. Indeed, for sampled distributions that differ only slightly from a Gaussian form, the accuracy of the test is remarkably good. Nevertheless, when applying the t-test, it is always important to remember that the assumption of a Gaussian parent population is central to the method.

31.7.6 Fisher's F-test

Having concentrated on tests for the mean μ of a Gaussian distribution, we now consider tests for its standard deviation σ. Before discussing Fisher's F-test for comparing the standard deviations of two samples, we begin by considering the case when an independent sample $x_1, x_2, \ldots, x_N$ is drawn from a Gaussian

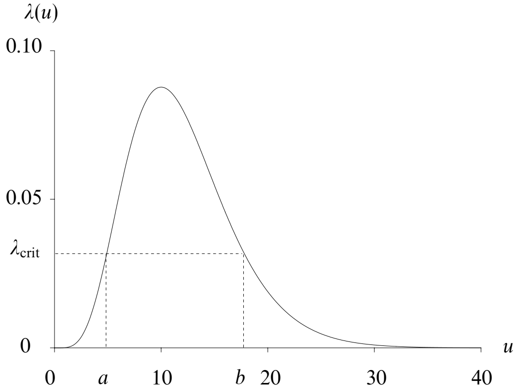


Figure 31.12 The sampling distribution $P(u|H_0)$ for $N = 10$; this is a chi-squared distribution for $N-1$ degrees of freedom.

distribution with unknown μ and σ, and we wish to distinguish between the two hypotheses

$$H_0 : \sigma^2 = \sigma_0^2, \quad -\infty < \mu < \infty \qquad \text{and} \qquad H_1 : \sigma^2 \neq \sigma_0^2, \quad -\infty < \mu < \infty,$$

where σ_0^2 is a given number. Here, the parameter space $\mathcal{A}$ is the half-plane $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$, whereas the subspace $\mathcal{S}$ characterised by the null hypothesis H_0 is the line $\sigma^2 = \sigma_0^2$, $-\infty < \mu < \infty$.

The likelihood function for this situation is given by

$$L(\mathbf{x}; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{\sum_i (x_i - \mu)^2}{2\sigma^2}\right].$$

The maximum of L in $\mathcal{A}$ occurs at $\mu = \bar{x}$ and $\sigma^2 = s^2$, whereas the maximum of L in $\mathcal{S}$ is at $\mu = \bar{x}$ and $\sigma^2 = \sigma_0^2$. Thus, the generalised likelihood ratio is given by

$$\lambda(\mathbf{x}) = \frac{L(\mathbf{x}; \bar{x}, \sigma_0^2)}{L(\mathbf{x}; \bar{x}, s^2)} = \left(\frac{u}{N}\right)^{N/2} \exp\left[-\tfrac{1}{2}(u - N)\right],$$

where we have introduced the variable

$$u = \frac{Ns^2}{\sigma_0^2} = \frac{\sum_i (x_i - \bar{x})^2}{\sigma_0^2}. \tag{31.121}$$

An example of this distribution is plotted in figure 31.12 for $N = 10$. From the figure, we see that the rejection region $\lambda < \lambda_{\rm crit}$ corresponds to a two-tailed rejection region on u given by

$$0 < u < a \qquad \text{and} \qquad b < u < \infty,$$

where a and b are such that $\lambda_{\rm crit}(a) = \lambda_{\rm crit}(b)$, as shown in figure 31.12. In practice, however, it is difficult to determine a and b for a given significance level α, so a slightly different rejection region, which we now describe, is usually adopted.

The sampling distribution $P(u|H_0)$ may be found straightforwardly from the sampling distribution of s given in (31.35). Let us first determine $P(s^2|H_0)$ by demanding that

$$P(s|H_0)\,ds = P(s^2|H_0)\,d(s^2),$$

from which we find

$$P(s^2|H_0) = \frac{P(s|H_0)}{2s} = \left(\frac{N}{2\sigma_0^2}\right)^{(N-1)/2} \frac{(s^2)^{(N-3)/2}}{\Gamma\left(\frac{1}{2}(N-1)\right)} \exp\left(-\frac{Ns^2}{2\sigma_0^2}\right). \qquad (31.122)$$

Thus, the sampling distribution of $u = Ns^2/\sigma_0^2$ is given by

$$P(u|H_0) = \frac{1}{2^{(N-1)/2}\Gamma\left(\frac{1}{2}(N-1)\right)} u^{(N-3)/2} \exp\left(-\tfrac{1}{2}u\right).$$

We note, in passing, that the distribution of u is precisely that of an $(N-1)$th-order chi-squared variable (see subsection 30.9.4), i.e. $u \sim \chi^2_{N-1}$. Although it does not give quite the best test, one then takes the rejection region to be

$$0 < u < a \qquad \text{and} \qquad b < u < \infty,$$

with a and b chosen such that the two tails have *equal areas*; the advantage of this choice is that tabulations of the chi-squared distribution make the size of this region relatively easy to estimate. Thus, for a given significance level α, we have

$$\int_0^a P(u|H_0)\,du = \alpha/2 \qquad \text{and} \qquad \int_b^\infty P(u|H_0)\,du = \alpha/2.$$

►*Ten independent sample values x_i, $i = 1, 2, \ldots, 10$, are drawn at random from a Gaussian distribution with unknown mean μ and standard deviation σ. The sample values are as follows:*

2.22	2.56	1.07	0.24	0.18	0.95	0.73	−0.79	2.09	1.81

Test the null hypothesis $H_0 : \sigma^2 = 2$ at the 10% significance level.

For our null hypothesis $\sigma_0^2 = 2$. Since for this sample $s = 1.01$ and $N = 10$, from (31.121) we have $u = 5.10$. For $\alpha = 0.1$ we find, either numerically or using table 31.2, that $a = 3.33$ and $b = 16.92$. Thus, our rejection region is

$$0 < u < 3.33 \qquad \text{and} \qquad 16.92 < u < \infty.$$

The value $u = 5.10$ from our sample does not lie in the rejection region, and so we cannot reject the null hypothesis $H_0 : \sigma^2 = 2$. ◄

We now turn to Fisher's F-test. Let us suppose that two independent samples of sizes N_1 and N_2 are drawn from Gaussian distributions with means and variances μ_1, σ_1^2 and μ_2, σ_2^2 respectively, and we wish to distinguish between the two hypotheses

$$H_0 : \sigma_1^2 = \sigma_2^2 \qquad \text{and} \qquad H_1 : \sigma_1^2 \neq \sigma_2^2.$$

In this case, the generalised likelihood ratio is found to be

$$\lambda = \frac{(N_1+N_2)^{(N_1+N_2)/2}}{N_1^{N_1/2} N_2^{N_2/2}} \frac{\left[F(N_1-1)/(N_2-1)\right]^{N_1/2}}{\left[1+F(N_1-1)/(N_2-1)\right]^{(N_1+N_2)/2}},$$

where F is given by the variance ratio

$$F = \frac{N_1 s_1^2/(N_1-1)}{N_2 s_2^2/(N_2-1)} \equiv \frac{u^2}{v^2} \tag{31.123}$$

and s_1 and s_2 are the standard deviations of the two samples. On plotting λ as a function of F, it is apparent that the rejection region $\lambda < \lambda_{\text{crit}}$ corresponds to a two-tailed test on F. Nevertheless, as will shall see below, by defining the fraction (31.123) appropriately, it is customary to make a one-tailed test on F.

The distribution of F may be obtained in a reasonably straightforward manner by making use of the distribution of the sample variance s^2 given in (31.122). Under our null hypothesis H_0, the two Gaussian distributions share a common variance, which we denote by σ^2. Changing the variable in (31.122) from s^2 to u^2 we find that u^2 has the sampling distribution

$$P(u^2|H_0) = \left(\frac{N-1}{2\sigma^2}\right)^{(N-1)/2} \frac{1}{\Gamma\left(\frac{1}{2}(N-1)\right)} (u^2)^{(N-3)/2} \exp\left[-\frac{(N-1)u^2}{2\sigma^2}\right].$$

Since u^2 and v^2 are independent, their joint distribution is simply the product of their individual distributions and is given by

$$P(u^2|H_0)P(v^2|H_0) = A(u^2)^{(N_1-3)/2}(v^2)^{(N_2-3)/2} \exp\left[-\frac{(N_1-1)u^2+(N_2-1)v^2}{2\sigma^2}\right],$$

where the constant A is given by

$$A = \frac{(N_1-1)^{(N_1-1)/2}(N_2-1)^{(N_2-1)/2}}{2^{(N_1+N_2-2)/2}\sigma^{(N_1+N_2-2)}\Gamma\left(\frac{1}{2}(N_1-1)\right)\Gamma\left(\frac{1}{2}(N_2-1)\right)}. \tag{31.124}$$

Now, for fixed v we have $u^2 = Fv^2$ and $d(u^2) = v^2 dF$. Thus, the joint sampling distribution $P(v^2, F|H_0)$ is obtained by requiring that

$$P(v^2, F|H_0)\, d(v^2)\, dF = P(u^2|H_0)P(v^2|H_0)\, d(u^2)\, d(v^2). \tag{31.125}$$

In order to find the distribution of F alone, we now integrate $P(v^2, F|H_0)$ with respect to v^2 from 0 to ∞, from which we obtain

$$\begin{aligned} P(F|H_0) & \\ &= \left(\frac{N_1-1}{N_2-1}\right)^{(N_1-1)/2} \frac{F^{(N_1-3)/2}}{B\left(\frac{1}{2}(N_1-1), \frac{1}{2}(N_2-1)\right)} \left(1 + \frac{N_1-1}{N_2-1}F\right)^{-(N_1+N_2-2)/2}, \end{aligned} \tag{31.126}$$

where $B\left(\frac{1}{2}(N_1-1), \frac{1}{2}(N_2-1)\right)$ is the beta function defined in the Appendix. $P(F|H_0)$ is called the *F-distribution* (or occasionally the *Fisher distribution*) with $(N_1 - 1,\ N_2 - 1)$ degrees of freedom.

►*Evaluate the integral* $\int_0^\infty P(v^2, F|H_0)\, d(v^2)$ *to obtain result (31.126).*

From (31.125), we have

$$P(F|H_0) = AF^{(N_1-3)/2} \int_0^\infty (v^2)^{(N_1+N_2-4)/2} \exp\left\{-\frac{[(N_1-1)F + (N_2-1)]v^2}{2\sigma^2}\right\} d(v^2).$$

Making the substitution $x = [(N_1-1)F + (N_2-1)]v^2/(2\sigma^2)$, we obtain

$$\begin{aligned} P(F|H_0) &= A\left[\frac{2\sigma^2}{(N_1-1)F + (N_2-1)}\right]^{(N_1+N_2-2)/2} F^{(N_1-3)/2} \int_0^\infty x^{(N_1+N_2-4)/2} e^{-x}\, dx \\ &= A\left[\frac{2\sigma^2}{(N_1-1)F + (N_2-1)}\right]^{(N_1+N_2-2)/2} F^{(N_1-3)/2}\Gamma\left(\tfrac{1}{2}(N_1+N_2-2)\right), \end{aligned}$$

where in the last line we have used the definition of the gamma function given in the Appendix. Using the further result (18.165), which expresses the beta function in terms of the gamma function, and the expression for A given in (31.124), we see that $P(F|H_0)$ is indeed given by (31.126). ◄

As it does not matter whether the ratio F given in (31.123) is defined as u^2/v^2 or as v^2/u^2, it is conventional to put the larger sample variance on the top, so that F is always greater than or equal to unity. A large value of F indicates that the sample variances u^2 and v^2 are very different whereas a value of F close to unity means that they are very similar. Therefore, for a given significance α, it is

$C_{n_1,n_2}(F)$	$n_1 = 1$	2	3	4	5	6	7	8
$n_2 = 1$	161	200	216	225	230	234	237	239
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94
	$n_1 = 9$	10	20	30	40	50	100	∞
$n_2 = 1$	241	242	248	250	251	252	253	254
2	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
3	8.81	8.79	8.66	8.62	8.59	8.58	8.55	8.53
4	6.00	5.96	5.80	5.75	5.72	5.70	5.66	5.63
5	4.77	4.74	4.56	4.50	4.46	4.44	4.41	4.37
6	4.10	4.06	3.87	3.81	3.77	3.75	3.71	3.67
7	3.68	3.64	3.44	3.38	3.34	3.32	3.27	3.23
8	3.39	3.35	3.15	3.08	3.04	3.02	2.97	2.93
9	3.18	3.14	2.94	2.86	2.83	2.80	2.76	2.71
10	3.02	2.98	2.77	2.70	2.66	2.64	2.59	2.54
20	2.39	2.35	2.12	2.04	1.99	1.97	1.91	1.84
30	2.21	2.16	1.93	2.69	1.79	1.76	1.70	1.62
40	2.12	2.08	1.84	1.74	1.69	1.66	1.59	1.51
50	2.07	2.03	1.78	1.69	1.63	1.60	1.52	1.44
100	1.97	1.93	1.68	1.57	1.52	1.48	1.39	1.28
∞	1.88	1.83	1.57	1.46	1.39	1.35	1.24	1.00

Table 31.4 Values of F for which the cumulative probability function $C_{n_1,n_2}(F)$ of the F-distribution with (n_1, n_2) degrees of freedom has the value 0.95. For example, for $n_1 = 10$ and $n_2 = 6$, $C_{n_1,n_2}(4.06) = 0.95$.

customary to define the rejection region on F as $F > F_{\text{crit}}$, where

$$C_{n_1,n_2}(F_{\text{crit}}) = \int_1^{F_{\text{crit}}} P(F|H_0)\, dF = \alpha,$$

and $n_1 = N_1 - 1$ and $n_2 = N_2 - 1$ are the numbers of degrees of freedom. Table 31.4 lists values of F_{crit} corresponding to the 5% significance level (i.e. $\alpha = 0.05$) for various values of n_1 and n_2.

►*Suppose that two classes of students take the same mathematics examination and the following percentage marks are obtained:*

Class 1:	66	62	34	55	77	80	55	60	69	47	50
Class 2:	64	90	76	56	81	72	70				

Assuming that the two sets of examinations marks are drawn from Gaussian distributions, test the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ *at the 5% significance level.*

The variances of the two samples are $s_1^2 = (12.8)^2$ and $s_2^2 = (10.3)^2$ and the sample sizes are $N_1 = 11$ and $N_2 = 7$. Thus, we have

$$u^2 = \frac{N_1 s_1^2}{N_1 - 1} = 180.2 \quad \text{and} \quad v^2 = \frac{N_2 s_2^2}{N_2 - 1} = 123.8,$$

where we have taken u^2 to be the larger value. Thus, $F = u^2/v^2 = 1.46$ to two decimal places. Since the first sample contains eleven values and the second contains seven values, we take $n_1 = 10$ and $n_2 = 6$. Consulting table 31.4, we see that, at the 5% significance level, $F_{\rm crit} = 4.06$. Since our value lies comfortably below this, we conclude that there is no statistical evidence for rejecting the hypothesis that the two samples were drawn from Gaussian distributions with a common variance. ◄

It is also common to define the variable $z = \frac{1}{2} \ln F$, the distribution of which can be found straightfowardly from (31.126). This is a useful change of variable since it can be shown that, for large values of n_1 and n_2, the variable z is distributed approximately as a Gaussian with mean $\frac{1}{2}(n_2^{-1} - n_1^{-1})$ and variance $\frac{1}{2}(n_2^{-1} + n_1^{-1})$.

31.7.7 Goodness of fit in least-squares problems

We conclude our discussion of hypothesis testing with an example of a goodness-of-fit test. In section 31.6, we discussed the use of the method of least squares in estimating the best-fit values of a set of parameters $\mathbf{a}$ in a given model $y = f(x; \mathbf{a})$ for a data set (x_i, y_i), $i = 1, 2, \ldots, N$. We have not addressed, however, the question of whether the best-fit model $y = f(x; \hat{\mathbf{a}})$ does, in fact, provide a good fit to the data. In other words, we have not considered thus far how to verify that the functional form f of our assumed model is indeed correct. In the language of hypothesis testing, we wish to distinguish between the two hypotheses

$$H_0 : \text{model is correct} \qquad \text{and} \qquad H_1 : \text{model is incorrect}.$$

Given the vague nature of the alternative hypothesis H_1, we clearly cannot use the generalised likelihood-ratio test. Nevertheless, it is still possible to test the null hypothesis H_0 at a given significance level α.

The least-squares estimates of the parameters $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_M$, as discussed in section 31.6, are those values that minimise the quantity

$$\chi^2(\mathbf{a}) = \sum_{i,j=1}^{N} [y_i - f(x_i; \mathbf{a})](\mathsf{N}^{-1})_{ij}[y_j - f(x_j; \mathbf{a})] = (\mathsf{y} - \mathsf{f})^{\rm T}\mathsf{N}^{-1}(\mathsf{y} - \mathsf{f}).$$

In the last equality, we rewrote the expression in matrix notation by defining the column vector f with elements $f_i = f(x_i;\mathbf{a})$. The value $\chi^2(\hat{\mathbf{a}})$ at this minimum can be used as a statistic to test the null hypothesis H_0, as follows. The N quantities $y_i - f(x_i;\mathbf{a})$ are Gaussian distributed. However, provided the function $f(x_j;\mathbf{a})$ is linear in the parameters $\mathbf{a}$, the equations (31.98) that determine the least-squares estimate $\hat{\mathbf{a}}$ constitute a set of M linear constraints on these N quantities. Thus, as discussed in subsection 30.15.2, the sampling distribution of the quantity $\chi^2(\hat{\mathbf{a}})$ will be a *chi-squared distribution with* $N-M$ *degrees of freedom* (d.o.f), which has the expectation value and variance

$$E[\chi^2(\hat{\mathbf{a}})] = N - M \qquad \text{and} \qquad V[\chi^2(\hat{\mathbf{a}})] = 2(N-M).$$

Thus we would expect the value of $\chi^2(\hat{\mathbf{a}})$ to lie typically in the range $(N-M) \pm \sqrt{2(N-M)}$. A value lying outside this range may suggest that the assumed model for the data is incorrect. A very small value of $\chi^2(\hat{\mathbf{a}})$ is usually an indication that the model has too many free parameters and has 'over-fitted' the data. More commonly, the assumed model is simply incorrect, and this usually results in a value of $\chi^2(\hat{\mathbf{a}})$ that is larger than expected.

One can choose to perform either a one-tailed or a two-tailed test on the value of $\chi^2(\hat{\mathbf{a}})$. It is usual, for a given significance level α, to define the one-tailed rejection region to be $\chi^2(\hat{\mathbf{a}}) > k$, where the constant k satisfies

$$\int_k^\infty P(\chi_n^2)\, d\chi_n^2 = \alpha \tag{31.127}$$

and $P(\chi_n^2)$ is the PDF of the chi-squared distribution with $n = N - M$ degrees of freedom (see subsection 30.9.4).

►*An experiment produces the following data sample pairs* (x_i, y_i):

x_i:	1.85	2.72	2.81	3.06	3.42	3.76	4.31	4.47	4.64	4.99
y_i:	2.26	3.10	3.80	4.11	4.74	4.31	5.24	4.03	5.69	6.57

where the x_i*-values are known exactly but each* y_i*-value is measured only to an accuracy of* $\sigma = 0.5$. *At the one-tailed* 5% *significance level, test the null hypothesis* H_0 *that the underlying model for the data is a straight line* $y = mx + c$.

These data are the same as those investigated in section 31.6 and plotted in figure 31.9. As shown previously, the least squares estimates of the slope m and intercept c are given by

$$\hat{m} = 1.11 \qquad \text{and} \qquad \hat{c} = 0.4. \tag{31.128}$$

Since the error on each y_i-value is drawn independently from a Gaussian distribution with standard deviation σ, we have

$$\chi^2(\mathbf{a}) = \sum_{i=1}^N \left[\frac{y_i - f(x_i;\mathbf{a})}{\sigma}\right]^2 = \sum_{i=1}^N \left[\frac{y_i - mx_i - c}{\sigma}\right]^2. \tag{31.129}$$

Inserting the values (31.128) into (31.129), we obtain $\chi^2(\hat{m},\hat{c}) = 11.5$. In our case, the number of data points is $N = 10$ and the number of fitted parameters is $M = 2$. Thus, the number of degrees of freedom is $n = N - M = 8$. Setting $n = 8$ and $\alpha = 0.05$ in (31.127) we find from table 31.2 that $k = 15.51$. Hence our rejection region is

$$\chi^2(\hat{m}, \hat{c}) > 15.51.$$

Since above we found $\chi^2(\hat{m}, \hat{c}) = 11.5$, we cannot reject the null hypothesis that the underlying model for the data is a straight line $y = mx + c$. ◄

As mentioned above, our analysis is only valid if the function $f(x; \mathbf{a})$ is linear in the parameters $\mathbf{a}$. Nevertheless, it is so convenient that it is sometimes applied in non-linear cases, provided the non-linearity is not too severe.

31.8 Exercises

31.1 A group of students uses a pendulum experiment to measure g, the acceleration of free fall, and obtains the following values (in m s^{-2}): 9.80, 9.84, 9.72, 9.74, 9.87, 9.77, 9.28, 9.86, 9.81, 9.79, 9.82. What would you give as the best value and standard error for g as measured by the group?

31.2 Measurements of a certain quantity gave the following values: 296, 316, 307, 278, 312, 317, 314, 307, 313, 306, 320, 309. Within what limits would you say there is a 50% chance that the correct value lies?

31.3 The following are the values obtained by a class of 14 students when measuring a physical quantity x: 53.8, 53.1, 56.9, 54.7, 58.2, 54.1, 56.4, 54.8, 57.3, 51.0, 55.1, 55.0, 54.2, 56.6.

(a) Display these results as a histogram and state what you would give as the best value for x.
(b) Without calculation, estimate how much reliance could be placed upon your answer to (a).
(c) Databooks give the value of x as 53.6 with negligible error. Are the data obtained by the students in conflict with this?

31.4 Two physical quantities x and y are connected by the equation

$$y^{1/2} = \frac{x}{ax^{1/2} + b},$$

and measured pairs of values for x and y are as follows:

x:	10	12	16	20
y:	409	196	114	94

Determine the best values for a and b by graphical means, and (either by hand or by using a built-in calculator routine) by a least-squares fit to an appropriate straight line.

31.5 Measured quantities x and y are known to be connected by the formula

$$y = \frac{ax}{x^2 + b},$$

where a and b are constants. Pairs of values obtained experimentally are

x:	2.0	3.0	4.0	5.0	6.0
y:	0.32	0.29	0.25	0.21	0.18

Use these data to make best estimates of the values of y that would be obtained for (a) $x = 7.0$, and (b) $x = -3.5$. As measured by fractional error, which estimate is likely to be the more accurate?

31.6 Prove that the sample mean is the best *linear unbiased estimator* of the population mean μ as follows.

(a) If the real numbers $a_1, a_2, \ldots, a_n$ satisfy the constraint $\sum_{i=1}^n a_i = C$, where C is a given constant, show that $\sum_{i=1}^n a_i^2$ is minimised by $a_i = C/n$ for all i.

(b) Consider the linear estimator $\hat{\mu} = \sum_{i=1}^n a_i x_i$. Impose the conditions (i) that it is *unbiased* and (ii) that it is as *efficient* as possible.

31.7 A population contains individuals of k types in equal proportions. A quantity X has mean μ_i amongst individuals of type i and variance σ^2, which has the same value for all types. In order to estimate the mean of X over the whole population, two schemes are considered; each involves a total sample size of nk. In the first the sample is drawn randomly from the whole population, whilst in the second (*stratified sampling*) n individuals are randomly selected from each of the k types.

Show that in both cases the estimate has expectation

$$\mu = \frac{1}{k}\sum_{i=1}^{k} \mu_i,$$

but that the variance of the first scheme exceeds that of the second by an amount

$$\frac{1}{k^2 n}\sum_{i=1}^{k} (\mu_i - \mu)^2.$$

31.8 Carry through the following proofs of statements made in subsections 31.5.2 and 31.5.3 about the ML estimators $\hat{\tau}$ and $\hat{\lambda}$.

(a) Find the expectation values of the ML estimators $\hat{\tau}$ and $\hat{\lambda}$ given, respectively, in (31.71) and (31.75). Hence verify equations (31.76), which show that, even though an ML estimator is unbiased, it does not follow that functions of it are also unbiased.

(b) Show that $E[\hat{\tau}^2] = (N+1)\tau^2/N$ and hence prove that $\hat{\tau}$ is a minimum-variance estimator of τ.

31.9 Each of a series of experiments consists of a large, but unknown, number n ($\gg 1$) of trials in each of which the probability of success p is the same, but also unknown. In the ith experiment, $i = 1, 2, \ldots, N$, the total number of successes is x_i ($\gg 1$). Determine the log-likelihood function.

Using Stirling's approximation to $\ln(n - x)$, show that

$$\frac{d \ln(n - x)}{dn} \approx \frac{1}{2(n - x)} + \ln(n - x),$$

and hence evaluate $\partial({}^nC_x)/\partial n$.

By finding the (coupled) equations determining the ML estimators $\hat{p}$ and $\hat{n}$, show that, to order n^{-1}, they must satisfy the simultaneous 'arithmetic' and 'geometric' mean constraints

$$\hat{n}\hat{p} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and} \quad (1 - \hat{p})^N = \prod_{i=1}^{N}\left(1 - \frac{x_i}{\hat{n}}\right).$$

31.10 This exercise is intended to illustrate the dangers of applying formalised estimator techniques to distributions that are not well behaved in a statistical sense.

The following are five sets of 10 values, all drawn from the same Cauchy distribution with parameter a.

(i)	4.81	−1.24	1.30	−0.23	2.98
	−1.13	−8.32	2.62	−0.79	−2.85
(ii)	0.07	1.54	0.38	−2.76	−8.82
	1.86	−4.75	4.81	1.14	−0.66
(iii)	0.72	4.57	0.86	−3.86	0.30
	−2.00	2.65	−17.44	−2.26	−8.83
(iv)	−0.15	202.76	−0.21	−0.58	−0.14
	0.36	0.44	3.36	−2.96	5.51
(v)	0.24	−3.33	−1.30	3.05	3.99
	1.59	−7.76	0.91	2.80	−6.46

Ignoring the fact that the Cauchy distribution does not have a finite variance (or even a formal mean), show that $\hat{a}$, the ML estimator of a, has to satisfy

$$s(\hat{a}) = \sum_{i=1}^{10} \frac{1}{1 + x_i^2/\hat{a}^2} = 5. \quad (*)$$

Using a programmable calculator, spreadsheet or computer, find the value of $\hat{a}$ that satisfies $(*)$ for each of the data sets and compare it with the value $a = 1.6$ used to generate the data. Form an opinion regarding the variance of the estimator.

Show further that if it is assumed that $(E[\hat{a}])^2 = E[\hat{a}^2]$, then $E[\hat{a}] = \nu_2^{1/2}$, where ν_2 is the second (central) moment of the distribution, which for the Cauchy distribution is infinite!

31.11 According to a particular theory, two dimensionless quantities X and Y have equal values. Nine measurements of X gave values of 22, 11, 19, 19, 14, 27, 8, 24 and 18, whilst seven measured values of Y were 11, 14, 17, 14, 19, 16 and 14. Assuming that the measurements of both quantities are Gaussian distributed with a common variance, are they consistent with the theory? An alternative theory predicts that $Y^2 = \pi^2 X$; are the data consistent with this proposal?

31.12 On a certain (testing) steeplechase course there are 12 fences to be jumped, and any horse that falls is not allowed to continue in the race. In a season of racing a total of 500 horses started the course and the following numbers fell at each fence:

Fence:	1	2	3	4	5	6	7	8	9	10	11	12
Falls:	62	75	49	29	33	25	30	17	19	11	15	12

Use this data to determine the overall probability of a horse's falling at a fence, and test the hypothesis that it is the same for all horses and fences as follows.

(a) Draw up a table of the expected number of falls at each fence on the basis of the hypothesis.

(b) Consider for each fence i the standardised variable

$$z_i = \frac{\text{estimated falls} - \text{actual falls}}{\text{standard deviation of estimated falls}},$$

and use it in an appropriate χ^2 test.

(c) Show that the data indicates that the odds against all fences being equally testing are about 40 to 1. Identify the fences that are significantly easier or harder than the average.

31.13 A similar technique to that employed in exercise 31.12 can be used to test correlations between characteristics of sampled data. To illustrate this consider the following problem.

During an investigation into possible links between mathematics and classical music, pupils at a school were asked whether they had preferences (a) between mathematics and english, and (b) between classical and pop music. The results are given below.

	Classical	None	Pop
Mathematics	23	13	14
None	17	17	36
English	30	10	40

By computing tables of expected numbers, based on the assumption that no correlations exist, and calculating the relevant values of χ^2, determine whether there is any evidence for

(a) a link between academic and musical tastes, and
(b) a claim that pupils either had preferences in both areas or had no preference.

You will need to consider the appropriate value for the number of degrees of freedom to use when applying the χ^2 test.

31.14 Three candidates X, Y and Z were standing for election to a vacant seat on their college's Student Committee. The members of the electorate (current first-year students, consisting of 150 men and 105 women) were each allowed to cross out the name of the candidate they least wished to be elected, the other two candidates then being credited with one vote each. The following data are known.

(a) X received 100 votes from men, whilst Y received 65 votes from women.
(b) Z received five more votes from men than X received from women.
(c) The total votes cast for X and Y were equal.

Analyse this data in such a way that a χ^2 test can be used to determine whether voting was other than random (i) amongst men and (ii) amongst women.

31.15 A particle detector consisting of a shielded scintillator is being tested by placing it near a particle source whose intensity can be controlled by the use of absorbers. It might register counts even in the absence of particles from the source because of the cosmic ray background.

The number of counts n registered in a fixed time interval as a function of the source strength s is given in as:

source strength s:	0	1	2	3	4	5	6
counts n:	6	11	20	42	44	62	61

At any given source strength, the number of counts is expected to be Poisson distributed with mean

$$n = a + bs,$$

where a and b are constants. Analyse the data for a fit to this relationship and obtain the best values for a and b together with their standard errors.

(a) How well is the cosmic ray background determined?
(b) What is the value of the correlation coefficient between a and b? Is this consistent with what would happen if the cosmic ray background were imagined to be negligible?
(c) Do the data fit the expected relationship well? Is there any evidence that the reported data 'are too good a fit'?

31.16 The function $y(x)$ is known to be a quadratic function of x. The following table gives the measured values and uncorrelated standard errors of y measured at various values of x (in which there is negligible error):

x	1	2	3	4	5
$y(x)$	3.5 ± 0.5	2.0 ± 0.5	3.0 ± 0.5	6.5 ± 1.0	10.5 ± 1.0

Construct the response matrix R using as basis functions 1, x, x^2. Calculate the matrix $\mathsf{R}^{\mathrm{T}}\mathsf{N}^{-1}\mathsf{R}$ and show that its inverse, the covariance matrix V, has the form

$$\mathsf{V} = \frac{1}{9184}\begin{pmatrix} 12\,592 & -9708 & 1580 \\ -9708 & 8413 & -1461 \\ 1580 & -1461 & 269 \end{pmatrix}.$$

Use this matrix to find the best values, and their uncertainties, for the coefficients of the quadratic form for $y(x)$.

31.17 The following are the values and standard errors of a physical quantity $f(\theta)$ measured at various values of θ (in which there is negligible error):

θ	0	$\pi/6$	$\pi/4$	$\pi/3$
$f(\theta)$	3.72 ± 0.2	1.98 ± 0.1	-0.06 ± 0.1	-2.05 ± 0.1

θ	$\pi/2$	$2\pi/3$	$3\pi/4$	π
$f(\theta)$	-2.83 ± 0.2	1.15 ± 0.1	3.99 ± 0.2	9.71 ± 0.4

Theory suggests that f should be of the form $a_1 + a_2 \cos\theta + a_3 \cos 2\theta$. Show that the normal equations for the coefficients a_i are

$$\begin{aligned} 481.3a_1 + 158.4a_2 - 43.8a_3 &= 284.7, \\ 158.4a_1 + 218.8a_2 + 62.1a_3 &= -31.1, \\ -43.8a_1 + 62.1a_2 + 131.3a_3 &= 368.4. \end{aligned}$$

(a) If you have matrix inversion routines available on a computer, determine the best values and variances for the coefficients a_i and the correlation between the coefficients a_1 and a_2.

(b) If you have only a calculator available, solve for the values using a Gauss–Seidel iteration and start from the approximate solution $a_1 = 2$, $a_2 = -2$, $a_3 = 4$.

31.18 Prove that the expression given for the Student's t-distribution in equation (31.118) is correctly normalised.

31.19 Verify that the F-distribution $P(F)$ given explicitly in equation (31.126) is symmetric between the two data samples, i.e. that it retains the same form but with N_1 and N_2 interchanged, if F is replaced by $F' = F^{-1}$. Symbolically, if $P'(F')$ is the distribution of F' and $P(F) = \eta(F, N_1, N_2)$, then $P'(F') = \eta(F', N_2, N_1)$.

31.20 It is claimed that the two following sets of values were obtained (a) by randomly drawing from a normal distribution that is $N(0, 1)$ and then (b) randomly assigning each reading to one of two sets A and B:

Set A:	−0.314	0.603	−0.551	−0.537	−0.160	−1.635	0.719
	0.610	0.482	−1.757	0.058			
Set B:	−0.691	1.515	−1.642	−1.736	1.224	1.423	1.165

Make tests, including t- and F-tests, to establish whether there is any evidence that either claims is, or both claims are, false.

31.9 Hints and answers

31.1 Note that the reading of 9.28 m s^{-2} is clearly in error, and should not be used in the calculation; 9.80 ± 0.02 m s^{-2}.

31.3 (a) 55.1. (b) Note that two thirds of the readings lie within ± 2 of the mean and that 14 readings are being used. This gives a standard error in the mean ≈ 0.6. (c) Student's t has a value of about 2.5 for 13 d.o.f. (degrees of freedom), and therefore it is likely at the 3% significance level that the data are in conflict with the accepted value.

31.5 Plot or calculate a least-squares fit of either x^2 versus x/y or xy versus y/x to obtain $a \approx 1.19$ and $b \approx 3.4$. (a) 0.16; (b) -0.27. Estimate (b) is the more accurate because, using the fact that $y(-x) = -y(x)$, it is effectively obtained by *inter*polation rather than *extra*polation.

31.7 Recall that, because of the equal proportions of each type, the *expected* numbers of each type in the first scheme is n. Show that the variance of the estimator for the second scheme is $\sigma^2/(kn)$. When calculating that for the first scheme, recall that $\overline{x_i^2} = \mu_i^2 + \sigma^2$ and note that μ_i^2 can be written as $(\mu_i - \mu + \mu)^2$.

31.9 The log-likelihood function is

$$\ln L = \sum_{i=1}^{N} \ln {}^nC_{x_i} + \sum_{i=1}^{N} x_i \ln p + \left(Nn - \sum_{i=1}^{N} x_i \right) \ln(1-p);$$

$$\frac{\partial({}^nC_x)}{\partial n} \approx \ln\left(\frac{n}{n-x}\right) - \frac{x}{2n(n-x)}.$$

Ignore the second term on the RHS of the above to obtain

$$\sum_{i=1}^{N} \ln\left(\frac{n}{n-x_i}\right) + N\ln(1-p) = 0.$$

31.11 $\bar{X} = 18.0 \pm 2.2$, $\bar{Y} = 15.0 \pm 1.1$. $\hat{\sigma} = 4.92$ giving $t = 1.21$ for 14 d.o.f., and is significant only at the 75% level. Thus there is no significant disagreement between the data and the theory. For the second theory, only the mean values can be tested as Y^2 will not be Gaussian distributed. The difference in the means is $\bar{Y}^2 - \pi^2\bar{X} = 47 \pm 36$ and is only significantly different from zero at the 82% level. Again the data is consistent with the proposed theory.

31.13 Consider how many entries may be chosen freely in the table if all row and column totals are to match the observed values. It should be clear that for an $m \times n$ table the number of degrees of freedom is $(m-1)(n-1)$.

(a) In order to make the fractions expressing each preference or lack of preference correct, the expected distribution, if there were no correlation, must be

	Classical	None	Pop
Mathematics	17.5	10	22.5
None	24.5	14	31.5
English	28	16	36

This gives a χ^2 of 12.3 for four d.o.f., making it less than 2% likely, that no correlation exists.

(b) The expected distribution, if there were no correlation, is

	Music preference	No music preference
Academic preference	104	26
No academic preference	56	14

This gives a χ^2 of 1.2 for one d.o.f and no evidence for the claim.

31.15 As the distribution at each value of s is Poisson, the best estimate of the measurement error is the square root of the number of counts, i.e. $\sqrt{n(s)}$. Linear regression gives $a = 4.3 \pm 2.1$ and $b = 10.06 \pm 0.94$.

(a) The cosmic ray background must be present, since $n(0) \neq 0$ but its value of about 4 is uncertain to within a factor 2.
(b) The correlation coefficient between a and b is -0.63. Yes; if a were reduced towards zero then b would have to be increased to compensate.
(c) Yes, $\chi^2 = 4.9$ for five d.o.f., which is almost exactly the 'expected' value, neither too good nor too bad.

31.17 $a_1 = 2.02 \pm 0.06,\ a_2 = -2.99 \pm 0.09,\ a_3 = 4.90 \pm 0.10;\ r_{12} = -0.60.$

31.19 Note that $|dF| = |dF'/F'^2|$ and write

$$1 + \frac{N_1 - 1}{(N_2 - 1)F'} \quad \text{as} \quad \left[\frac{N_1 - 1}{(N_2 - 1)F'}\right]\left[1 + \frac{(N_2 - 1)F'}{N_1 - 1}\right].$$

Index

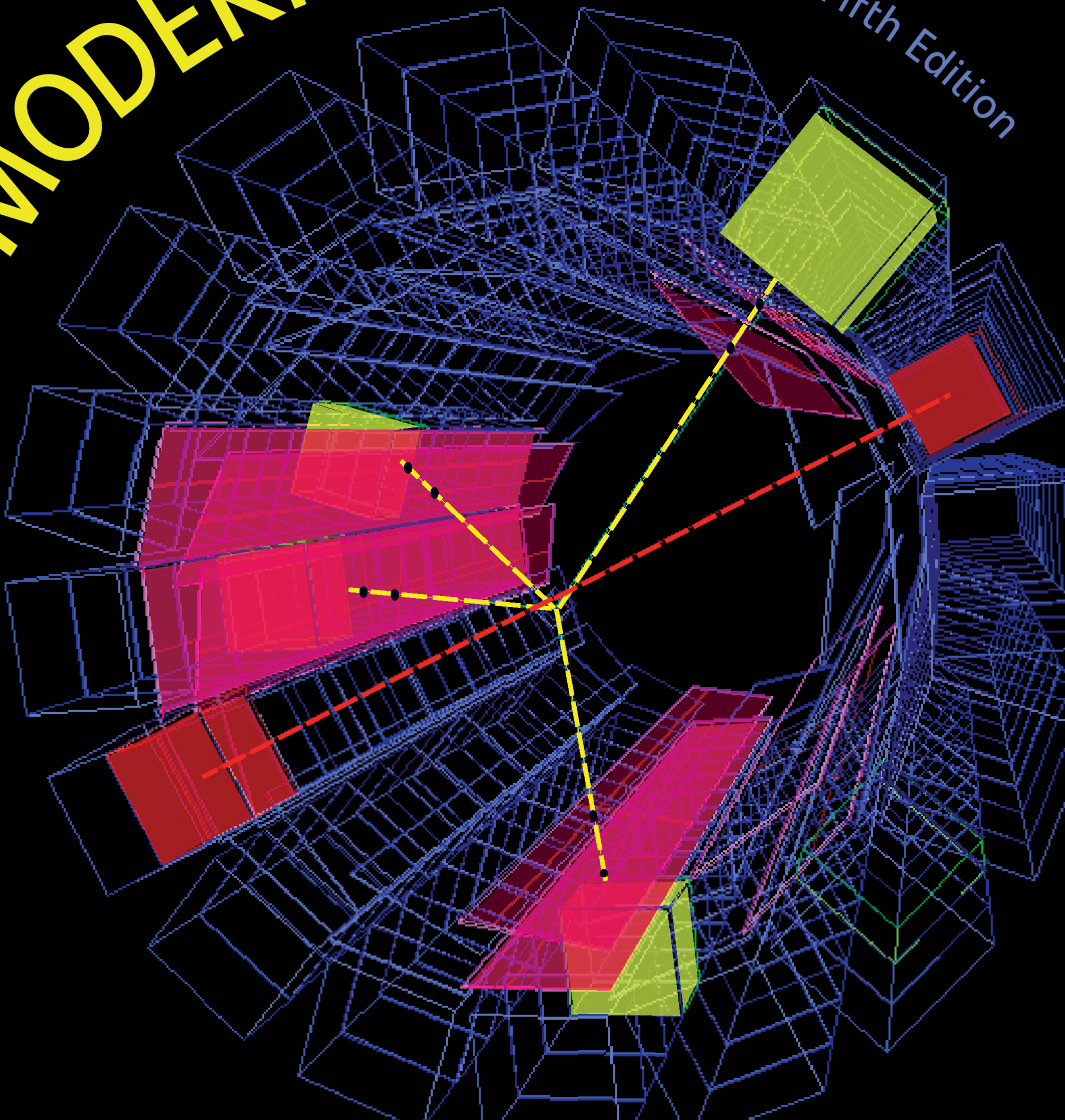
Where the discussion of a topic runs over two consecutive pages, reference is made only to the first of these. For discussions spread over three or more pages the first and last page numbers are given; these references are usually to the major treatment of the corresponding topic. Isolated refences to a topic, including those appearing on consecutive pages, are listed individually. Some long topics are split, e.g. 'Fourier transforms' and 'Fourier transforms, examples'. The letter 'n' after a page number indicates that the topic is discussed in a footnote on the relevant page.

3434

Paul A. Tipler
Ralph A. Llewellyn

MODERN PHYSICS

Fifth Edition



Publisher: Clancy Marshall
Senior Acquisitions Editor: Jessica Fiorillo
Marketing Manager: Anthony Palmiotto
Media Editors: Jeanette Picerno and Samantha Calamari
Supplements Editor and Editorial Assistant: Janie Chan
Senior Project Editor: Mary Louise Byrd
Cover and Text Designer: Diana Blume
Photo Editor: Ted Szczepanski
Photo Researcher: Rae Grant
Senior Illustration Coordinator: Bill Page
Production Coordinator: Paul W. Rohloff
Illustrations and Composition: Preparé
Printing and Binding: Quebecor Printing

Library of Congress Control Number: 2007931523

ISBN-13: 978-0-7167-7550-8
ISBN-10: 0-7167-7550-6

Printed in the United States of America

First printing

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
Houndmills, Basingstoke RG21 6XS, England
www.whfreeman.com

MODERN PHYSICS

Fifth Edition

Paul A. Tipler

Formerly of Oakland University

Ralph A. Llewellyn

University of Central Florida

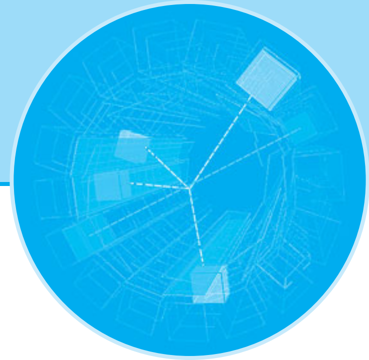
W. H. Freeman and Company • New York

3441

This page intentionally left blank

Contents

The indicates material that appears only on the Web site: www.whfreeman.com/tiplermodernphysics5e.
The indicates material of high interest to students.



Preface

In preparing this new edition of *Modern Physics,* we have again relied heavily on the many helpful suggestions from a large team of reviewers and from a host of instructor and student users of the earlier editions. Their advice reflected the discoveries that have further enlarged modern physics in the early years of this new century and took note of the evolution that is occurring in the teaching of physics in colleges and universities. As the term *modern physics* has come to mean the physics of the modern era—relativity and quantum theory—we have heeded the advice of many users and reviewers and preserved the historical and cultural flavor of the book while being careful to maintain the mathematical level of the fourth edition. We continue to provide the flexibility for instructors to match the book and its supporting ancillaries to a wide variety of teaching modes, including both one- and two-semester courses and media-enhanced courses.

Features

The successful features of the fourth edition have been retained, including the following:

- The logical structure—beginning with an introduction to relativity and quantization and following with applications—has been continued. Opening the book with relativity has been endorsed by many reviewers and instructors.
- As in the earlier editions, the end-of-chapter problems are separated into three sets based on difficulty, with the least difficult also grouped by chapter section. More than 10 percent of the problems in the fifth edition are new. The first edition's *Instructor's Solutions Manual* (*ISM*) with solutions, not just answers, to all end-of-chapter problems was the first such aid to accompany a physics (and not just a modern physics) textbook, and that leadership has been continued in this edition. The ISM is available in print or on CD for those adopting *Modern Physics,* fifth edition, for their classes. As with the previous edition, a paperback *Student's Solution Manual* containing one-quarter of the solutions in the ISM is also available.
- We have continued to include many examples in every chapter, a feature singled out by many instructors as a strength of the book. As before, we frequently use combined quantities such as hc, $\hbar c$, and ke^2 in eV • nm to simplify many numerical calculations.
- The summaries and reference lists at the end of every chapter have, of course, been retained and augmented, including the two-column format of the summaries, which improves their clarity.

- We have continued the use of real data in figures, photos of real people and apparatus, and short quotations from many scientists who were key participants in the development of modern physics. These features, along with the Notes at the end of each chapter, bring to life many events in the history of science and help counter the too-prevalent view among students that physics is a dull, impersonal collection of facts and formulas.
- More than two dozen Exploring sections, identified by an atom icon and dealing with text-related topics that captivate student interest such as superluminal speed and giant atoms, are distributed throughout the text.
- The book's Web site includes 30 MORE sections, which expand in depth on many text-related topics. These have been enthusiastically endorsed by both students and instructors and often serve as springboards for projects and alternate credit assignments. Identified by a laptop icon , each is introduced with a brief text box.
- More than 125 questions intended to foster discussion and review of concepts are distributed throughout the book. These have received numerous positive comments from many instructors over the years, often citing how the questions encourage deeper thought about the topic.
- Continued in the new edition are the Application Notes. These brief notes in the margins of many pages point to a few of the many benefits to society that have been made possible by a discovery or development in modern physics.

New Features

A number of new features are introduced in the fifth edition:

- The "Astrophysics and Cosmology" chapter that was on the fourth edition's Web site has been extensively rewritten and moved into the book as a new Chapter 13. Emphasis has been placed on presenting scientists' current understanding of the evolution of the cosmos based on the research in this dynamic field.
- The "Particle Physics" chapter has been substantially reorganized and rewritten focused on the remarkably successful Standard Model. As the new Chapter 12, it immediately precedes the new "Astrophysics and Cosmology" chapter to recognize the growing links between these active areas of current physics research.
- The two chapters concerned with the theory and applications of nuclear physics have been integrated into a new Chapter 11, "Nuclear Physics." Because of the renewed interest in nuclear power, that material in the fourth edition has been augmented and moved to a MORE section of the Web.
- Recognizing the need for students on occasion to be able to quickly review key concepts from classical physics that relate to topics developed in modern physics, we have added a new Classical Concept Review (CCR) to the book's Web site. Identified by a laptop icon in the margin near the pertinent modern physics topic of discussion, the CCR can be printed out to provide a convenient study support booklet.
- The *Instructor's Resource* CD for the fifth edition contains all the illustrations from the book in both PowerPoint and JPEG format. Also included is a gallery of the astronomical images from Chapter 13 in the original image colors.
- Several new MORE sections have been added to the book's Web site, and a few for which interest has waned have been removed.

Organization and Coverage

This edition, like the earlier ones, is divided into two parts: Part 1, "Relativity and Quantum Mechanics: The Foundation of Modern Physics," and Part 2, "Applications." We continue to open Part 1 with the two relativity chapters. This location for relativity is firmly endorsed by users and reviewers. The rationale is that this arrangement avoids separation of the foundations of quantum mechanics in Chapters 3 through 8 from its applications in Chapters 9 through 12. The two-chapter format for relativity provides instructors with the flexibility to cover only the basic concepts or to go deeper into the subject. Chapter 1 covers the essentials of special relativity and includes discussions of several paradoxes, such as the twin paradox and the pole-in-the-barn paradox, that never fail to excite student interest. Relativistic energy and momentum are covered in Chapter 2, which concludes with a mostly qualitative section on general relativity that emphasizes experimental tests. Because the relation $E^2 = p^2c^2 + (mc^2)^2$ is the result most needed for the later applications chapters, it is possible to omit Chapter 2 without disturbing continuity. Chapters 1 through 8 have been updated with a number of improved explanations and new diagrams. Several classical foundation topics in those chapters have been moved to the Classical Concept Review or recast as MORE sections. Many quantitative topics are included as MORE sections on the Web site. Examples of these are the derivation of Compton's equation (Chapter 3), the details of Rutherford's alpha-scattering theory (Chapter 4), the graphical solution of the finite square well (Chapter 6), and the excited states and spectra of two-electron atoms (Chapter 7). The comparisons of classical and quantum statistics are illustrated with several examples in Chapter 8, and unlike the other chapters in Part 1, Chapter 8 is arranged to be covered briefly and qualitatively if desired. This chapter, like Chapter 2, is not essential to the understanding of the applications chapters of Part 2 and may be used as an applications chapter or omitted without loss of continuity.

Preserving the approach used in the previous edition, in Part 2 the ideas and methods discussed in Part 1 are applied to the study of molecules, solids, nuclei, particles, and the cosmos. Chapter 9 ("Molecular Structure and Spectra") is a broad, detailed discussion of molecular bonding and the basic types of lasers. Chapter 10 ("Solid-State Physics") includes sections on bonding in metals, magnetism, and superconductivity. Chapter 11 ("Nuclear Physics") is an integration of the nuclear theory and applications that formed two chapters in the fourth edition. It focuses on nuclear structure and properties, radioactivity, and the applications of nuclear reactions. Included in the last topic are fission, fusion, and several techniques of age dating and elemental analysis. The material on nuclear power has been moved to a MORE section, and the discussion of radiation dosage continues as a MORE section. As mentioned above, Chapter 12 ("Particle Physics") has been substantially reorganized and rewritten with a focus on the Standard Model and revised to reflect the advances in that field since the earlier editions. The emphasis is on the fundamental interactions of the quarks, leptons, and force carriers and includes discussions of the conservation laws, neutrino oscillations, and supersymmetry. Finally, the thoroughly revised Chapter 13 ("Astrophysics and Cosmology") examines the current observations of stars and galaxies and qualitatively integrates our discussions of quantum mechanics, atoms, nuclei, particles, and relativity to explain our present understanding of the origin and evolution of the universe from the Big Bang to dark energy.

The Research Frontier

Research over the past century has added abundantly to our understanding of our world, forged strong links from physics to virtually every other discipline, and measurably improved the tools and devices that enrich life. As was the case at the beginning of the last century, it is hard for us to foresee in the early years of this century how scientific research will deepen our understanding of the physical universe and enhance the quality of life. Here are just a few of the current subjects of frontier research included in *Modern Physics,* fifth edition, that you will hear more of in the years just ahead. Beyond these years there will be many other discoveries that no one has yet dreamed of.

- **The Higgs boson**, the harbinger of mass, may now be within our reach at Brookhaven's Relativistic Heavy Ion Collider and at CERN with completion of the Large Hadron Collider. (Chapter 12)
- **The neutrino mass question** has been solved by the discovery of neutrino oscillations at the Super-Kamiokande and SNO neutrino observatories (Chapters 2, 11, and 12), but the magnitudes of the masses and whether the neutrino is a Majorana particle remain unanswered.
- **The origin of the proton's spin**, which may include contributions from virtual strange quarks, still remains uncertain. (Chapter 11)
- **The Bose-Einstein condensates**, which suggest atomic lasers and super–atomic clocks are in our future, were joined in 2003 by **Fermi-Dirac condensates,** wherein pairs of fermions act like bosons at very low temperatures. (Chapter 8)
- **It is now clear that dark energy** accounts for 74 percent of the mass/energy of the universe. Only 4 percent is baryonic (visible) matter. The remaining 22 percent consists of as yet unidentified **dark matter** particles. (Chapter 13)
- **The predicted fundamental particles of supersymmetry (SUSY)**, an integral part of grand unification theories, will be a priority search at the Large Hadron Collider. (Chapters 12 and 13)
- **High-temperature superconductors reached critical temperatures greater than 130 K a few years ago and doped fullerenes compete with cuprates for high-T_c records,** but a theoretical explanation of the phenomenon is not yet in hand. (Chapter 10)
- **Gravity waves from space** may soon be detected by the upgraded Laser Interferometric Gravitational Observatory (LIGO) and several similar laboratories around the world. (Chapter 2)
- **Adaptive-optics telescopes, large baseline arrays, and the Hubble telescope** are providing new views deeper into space of the very young universe, revealing that the expansion is speeding up, a discovery supported by results from the Sloan Digital Sky Survey and the Wilkinson Microwave Anisotropy Project. (Chapter 13)
- **Giant Rydberg atoms**, made accessible by research on tunable dye lasers, are now of high interest and may provide the first direct test of the correspondence principle. (Chapter 4)
- **The search for new elements has reached $Z = 118$,** tantalizingly near the edge of the "island of stability." (Chapter 11)

Many more discoveries and developments just as exciting as these are to be found throughout *Modern Physics,* fifth edition.

Some Teaching Suggestions

This book is designed to serve well in either one- or two-semester courses. The chapters in Part 2 are independent of one another and can be covered in any order. Some possible one-semester courses might consist of

- Part 1, Chapters 1, 3, 4, 5, 6, 7; and Part 2, Chapters 11, 12
- Part 1, Chapters 3, 4, 5, 6, 7, 8; and Part 2, Chapters 9, 10
- Part 1, Chapters 1, 2, 3, 4, 5, 6, 7; and Part 2, Chapter 9
- Part 1, Chapters 1, 3, 4, 5, 6, 7; and Part 2, Chapters 11, 12, 13

Possible two-semester courses might consist of

- Part 1, Chapters 1, 3, 4, 5, 6, 7; and Part 2, Chapters 9, 10, 11, 12, 13
- Part 1, Chapters 1, 2, 3, 4, 5, 6, 7, 8; and Part 2, Chapters 9, 10, 11, 12, 13

There is tremendous potential for individual student projects and alternate credit assignments based on the Exploring and, in particular, the MORE sections. The latter will encourage students to search for related sources on the Web.

Acknowledgments

Many people contributed to the success of the earlier editions of this book, and many more have helped with the development of the fifth edition. We owe our thanks to them all. Those who reviewed all or parts of this book, offering suggestions for the fifth edition, include

Marco Battaglia
University of California–Berkeley

Mario Belloni
Davidson College

Eric D. Carlson
Wake Forest University

David Cinabro
Wayne State University

Carlo Dallapiccola
University of Massachusetts–Amherst

Anthony D. Dinsmore
University of Massachusetts–Amherst

Ian T. Durham
Saint Anselm College

Jason J. Engbrecht
St. Olaf College

Brian Fick
Michigan Technological University

Massimiliano Galeazzi
University of Miami

Hugh Gallagher
Tufts University

Richard Gelderman
Western Kentucky University

Tim Gfroerer
Davidson College

Torgny Gustafsson
Rutgers University

Scott Heinekamp
Wells College

Adrian Hightower
Occidental College

Mark Hollabaugh
Normandale Community College

Richard D. Holland II
Southern Illinois University at Carbondale

Bei-Lok Hu
University of Maryland–College Park

Dave Kieda
University of Utah

Steve Kraemer
Catholic University of America

Wolfgang Lorenzon
University of Michigan

Bryan A. Luther
Concordia College at Moorhead

Catherine Mader
Hope College

Kingshuk Majumdar
Berea College

Peter Moeck
Portland State University

Robert M. Morse
University of Wisconsin–Madison

Igor Ostrovskii
University of Mississippi at Oxford

Anne Reilly
College of William and Mary

David Reitze
University of Florida

Mark Riley
Florida State University

Nitin Samarth
Pennsylvania State University

Kate Scholberg
Duke University

Ben E. K. Sugerman
Goucher College

Rein Uritam
Physics Department
Boston College

Ken Voss
University of Miami

Thad Walker
University of Wisconsin–Madison

Barry C. Walker
University of Delaware

Eric Wells
Augustana College

William R. Wharton
Wheaton College

Weldon J. Wilson
University of Central Oklahoma

R. W. M. Woodside
University College of Fraser Valley

We also thank the reviewers of the fourth and third editions. Their comments significantly influenced and shaped the fifth edition as well. For the fourth edition they were Darin Acosta, University of Florida; Jeeva Anandan, University of South Carolina; Gordon Aubrecht, Ohio State University; David A. Bahr, Bemidji State University; Patricia C. Boeshaar, Drew University; David P. Carico, California Polytechnic State University at San Luis Obispo; David Church, University of Washington; Wei Cui, Purdue University; Snezana Dalafave, College of New Jersey; Richard Gass, University of Cincinnati; David Gerdes, University of Michigan; Mark Hollabaugh, Normandale Community College; John L. Hubisz, North Carolina State University; Ronald E. Jodoin, Rochester Institute of Technology; Edward R. Kinney, University of Colorado at Boulder; Paul D. Lane, University of St. Thomas; Fernando J. Lopez-Lopez, Southwestern College; Dan MacIsaac, Northern Arizona University; Robert Pompi, SUNY at Binghamton; Warren Rogers, Westmont College; George Rutherford, Illinois State University; Nitin Samarth, Pennsylvania State University; Martin A. Sanzari, Fordham University; Earl E. Scime, West Virginia University; Gil Shapiro, University of California at Berkeley; Larry Solanch, Georgia College & State University; Francis M. Tam, Frostburg State University; Paul Tipton, University of Rochester; K. Thad Walker, University of Wisconsin at Madison; Edward A. Whittaker, Stevens Institute of Technology; Stephen Yerian, Xavier University; and Dean Zollman, Kansas State University.

For the third edition, reviewers were Bill Bassichis, Texas A&M University; Brent Benson, Lehigh University; H. J. Biritz, Georgia Institute of Technology; Patrick Briggs, The Citadel; David A. Briodo, Boston College; Tony Buffa, California Polytechnic State University at San Luis Obispo; Duane Carmony, Purdue University; Ataur R. Chowdhury, University of Alaska at Fairbanks; Bill Fadner, University of Northern Colorado; Ron Gautreau, New Jersey Institute of Technology; Charles Glashauser, Rutgers–The State University of New Jersey; Roger Hanson, University of Northern Iowa; Gary G. Ihas, University of Florida; Yuichi Kubota, University of Minnesota; David Lamp, Texas Tech University; Philip Lippel, University of Texas at Arlington; A. E. Livingston, University of Notre Dame; Steve Meloma, Gustavus Adolphus College; Benedict Y. Oh, Pennsylvania State University; Paul Sokol, Pennsylvania State University; Thor F. Stromberg, New Mexico State University; Maurice Webb, University of Wisconsin at Madison; and Jesse Weil, University of Kentucky.

All offered valuable suggestions for improvements, and we appreciate their help.

In addition, we give a special thanks to all the physicists and students from around the world who took time to send us kind words about the third and fourth editions and offered suggestions for improvements.

Finally, though certainly not least, we are grateful for the support, encouragement, and patience of our families throughout the project. We especially want to thank Mark Llewellyn for his preparation of the *Instructor's Solutions Manual* and the *Student's Solutions Manual* and for his numerous helpful suggestions from the very beginning of the project, Eric Llewellyn for his photographic and computer-generated images, David Jonsson at Uppsala University for his critical reading of every chapter of the fourth edition, and Jeanette Picerno for her imaginative work on the Web site. Finally, to the entire *Modern Physics* team at W. H. Freeman and Company goes our sincerest appreciation for their skill, hard work, understanding about deadlines, and support in bringing it all together.

Paul A. Tipler, *Berkeley, California*

Ralph A. Llewellyn, *Oviedo, Florida*

This page intentionally left blank

PART 1

Relativity and Quantum Mechanics: The Foundations of Modern Physics

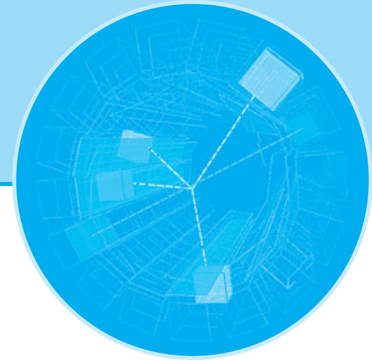
The earliest recorded systematic efforts to assemble knowledge about motion as a key to understanding natural phenomena were those of the ancient Greeks. Set forth in sophisticated form by Aristotle, theirs was a natural philosophy (i.e., physics) of explanations deduced from assumptions rather than experimentation. For example, it was a fundamental assumption that every substance had a "natural place" in the universe. Motion then resulted when a substance was trying to reach its natural place. Time was given a similar absolute meaning, as moving from some instant in the past (the creation of the universe) toward some end goal in the future, its natural place. The remarkable agreement between the deductions of Aristotelian physics and motions observed throughout the physical universe, together with a nearly total absence of accurate instruments to make contradictory measurements, led to acceptance of the Greek view for nearly 2000 years. Toward the end of that time a few scholars had begun to deliberately test some of the predictions of theory, but it was Italian scientist Galileo Galilei who, with his brilliant experiments on motion, established for all time the absolute necessity of experimentation in physics and, coincidentally, initiated the disintegration of Aristotelian physics. Within 100 years Isaac Newton had generalized the results of Galileo's experiments into his three spectacularly successful laws of motion, and the natural philosophy of Aristotle was gone.

With the burgeoning of experimentation, the following 200 years saw a multitude of major discoveries and a concomitant development of physical theories to explain them. Most of the latter, then as now, failed to survive increasingly sophisticated experimental tests, but by the dawn of the twentieth century Newton's theoretical explanation of the motion of mechanical systems had been joined by equally impressive laws of electromagnetism and thermodynamics as expressed by Maxwell, Carnot, and others. The remarkable success of these laws led many scientists to believe that description of the physical universe was complete. Indeed, A. A. Michelson, speaking to scientists near the end of the nineteenth century, said, "The grand underlying principles have been firmly established . . . the future truths of physics are to be looked for in the sixth place of decimals."

Such optimism (or pessimism, depending on your point of view) turned out to be premature, as there were already vexing cracks in the foundation of what we now refer to as classical physics. Two of these were described by Lord Kelvin, in his famous Baltimore Lectures in 1900, as the "two clouds" on the horizon of twentieth-century physics: the failure of theory to account for the radiation spectrum emitted by a blackbody and the inexplicable results of the Michelson-Morley experiment. Indeed, the breakdown of classical physics occurred in many different areas: the Michelson-Morley null result contradicted Newtonian relativity, the blackbody radiation spectrum contradicted predictions of thermodynamics, the photoelectric effect and the spectra of atoms could not be explained by electromagnetic theory, and the exciting discoveries of x rays and radioactivity seemed to be outside the framework of classical physics entirely. The development of the theories of quantum mechanics and relativity in the early twentieth century not only dispelled Kelvin's "dark clouds," they provided answers to all of the puzzles listed here and many more. The applications of these theories to such microscopic systems as atoms, molecules, nuclei, and fundamental particles and to macroscopic systems of solids, liquids, gases, and plasmas have given us a deep understanding of the intricate workings of nature and have revolutionized our way of life.

In Part 1 we discuss the foundations of the physics of the modern era, relativity theory, and quantum mechanics. Chapter 1 examines the apparent conflict between Einstein's principle of relativity and the observed constancy of the speed of light and shows how accepting the validity of both ideas led to the special theory of relativity. Chapter 2 discusses the relations connecting mass, energy, and momentum in special relativity and concludes with a brief discussion of general relativity and some experimental tests of its predictions. In Chapters 3, 4, and 5 the development of quantum theory is traced from the earliest evidences of quantization to de Broglie's hypothesis of electron waves. An elementary discussion of the Schrödinger equation is provided in Chapter 6, illustrated with applications to one-dimensional systems. Chapter 7 extends the application of quantum mechanics to many-particle systems and introduces the important new concepts of electron spin and the exclusion principle. Concluding the development, Chapter 8 discusses the wave mechanics of systems of large numbers of identical particles, underscoring the importance of the symmetry of wave functions. Beginning with Chapter 3, the chapters in Part 1 should be studied in sequence because each of Chapters 4 through 8 depends on the discussions, developments, and examples of the previous chapters.

CHAPTER 1



Relativity I

The relativistic character of the laws of physics began to be apparent very early in the evolution of classical physics. Even before the time of Galileo and Newton, Nicolaus Copernicus[1] had shown that the complicated and imprecise Aristotelian method of computing the motions of the planets, based on the assumption that Earth was located at the center of the universe, could be made much simpler, though no more accurate, if it were assumed that the planets move about the Sun instead of Earth. Although Copernicus did not publish his work until very late in life, it became widely known through correspondence with his contemporaries and helped pave the way for acceptance a century later of the heliocentric theory of planetary motion. While the Copernican theory led to a dramatic revolution in human thought, the aspect that concerns us here is that it did not consider the location of Earth to be special or favored in any way. Thus, the laws of physics discovered on Earth could apply equally well with any point taken as the center—i.e., the same equations would be obtained regardless of the origin of coordinates. This invariance of the equations that express the laws of physics is what we mean by the term *relativity*.

We will begin this chapter by investigating briefly the relativity of Newton's laws and then concentrate on the theory of relativity as developed by Albert Einstein (1879–1955). The theory of relativity consists of two rather different theories, the special theory and the general theory. The special theory, developed by Einstein and others in 1905, concerns the comparison of measurements made in different frames of reference moving with constant velocity relative to each other. Contrary to popular opinion, the special theory is not difficult to understand. Its consequences, which can be derived with a minimum of mathematics, are applicable in a wide variety of situations in physics and engineering. On the other hand, the general theory, also developed by Einstein (around 1916), is concerned with accelerated reference frames and gravity. Although a thorough understanding of the general theory requires more sophisticated mathematics (e.g., tensor analysis), a number of its basic ideas and important predictions can be discussed at the level of this book. The general theory is of great importance in cosmology and in understanding events that occur in the vicinity of very large masses (e.g., stars) but is rarely encountered in other areas of physics and engineering. We will devote this chapter entirely to the special theory (often referred to as *special relativity*) and discuss the general theory in the final section of Chapter 2, following the sections concerned with special relativistic mechanics.

1-1 The Experimental Basis of Relativity

Classical Relativity

In 1687, with the publication of the *Philosophiae Naturalis Principia Mathematica,* Newton became the first person to generalize the observations of Galileo and others into the laws of motion that occupied much of your attention in introductory physics. The second of Newton's three laws is

$$\mathbf{F} = m\frac{d\mathbf{v}}{dt} = m\mathbf{a} \qquad \textbf{1-1}$$

where $d\mathbf{v}/dt = \mathbf{a}$ is the acceleration of the mass m when acted upon by a net force $\mathbf{F}$. Equation 1-1 also includes the first law, the *law of inertia*, by implication: if $\mathbf{F} = 0$, then $d\mathbf{v}/dt = 0$ also, i.e., $\mathbf{a} = 0$. (Recall that letters and symbols in boldface type are vectors.)

As it turns out, Newton's laws of motion only work correctly in *inertial reference frames*, that is, reference frames in which the law of inertia holds.[2] They also have the remarkable property that they are *invariant*, or unchanged, in any reference frame that moves with constant velocity relative to an inertial frame. Thus, all inertial frames are equivalent—there is no special or favored inertial frame relative to which absolute measurements of space and time could be made. Two such inertial frames are illustrated in Figure 1-1, arranged so that corresponding axes in S and S' are parallel and S' moves in the $+x$ direction at velocity $\mathbf{v}$ for an observer in S (or S moves in the $-x'$

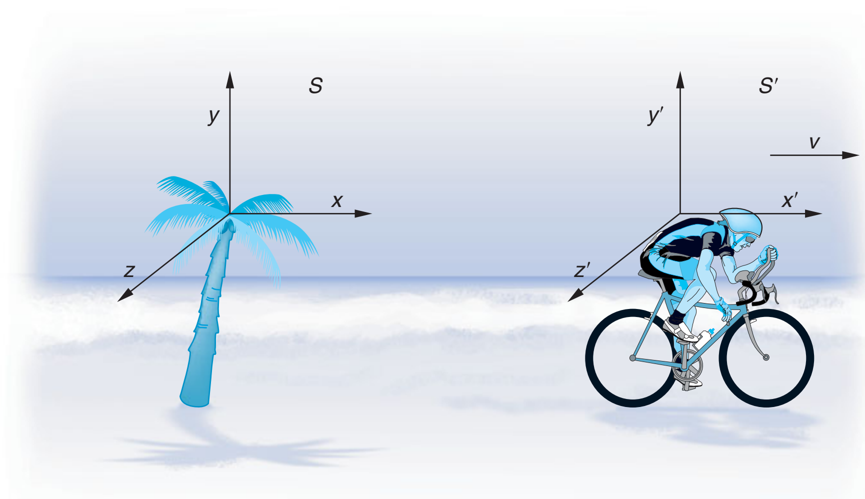


Figure 1-1 Inertial reference frame S is attached to Earth (the palm tree) and S' to the cyclist. The corresponding axes of the frames are parallel, and S' moves at speed v in the $+x$ direction of S.

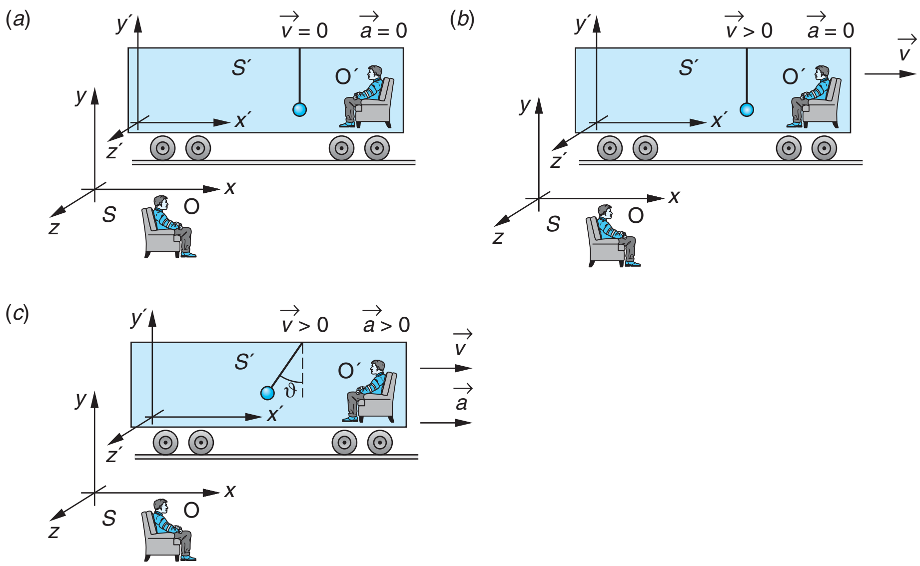


Figure 1-2 A mass suspended by a cord from the roof of a railroad boxcar illustrates the relativity of Newton's second law, $\mathbf{F} = m\mathbf{a}$. The only forces acting on the mass are its weight $m\mathbf{g}$ and the tension $\mathbf{T}$ in the cord. (*a*) The boxcar sits at rest in S. Since the velocity $\mathbf{v}$ and the acceleration $\mathbf{a}$ of the boxcar (i.e., the system S') are both zero, both observers see the mass hanging vertically at rest with $\mathbf{F} = \mathbf{F}' = 0$. (*b*) As S' moves in the $+x$ direction with $\mathbf{v}$ constant, both observers see the mass hanging vertically but moving at $\mathbf{v}$ with respect to O in S and at rest with respect to the S' observer. Thus, $\mathbf{F} = \mathbf{F}' = 0$. (*c*) As S' moves in the $+x$ direction with $\mathbf{a} > 0$ with respect to S, the mass hangs at an angle $\vartheta > 0$ with respect to the vertical. However, it is still at rest (i.e., in equilibrium) with respect to the observer in S', who now "explains" the angle ϑ by adding a pseudoforce $\mathbf{F}_\mathbf{p}$ in the $-x'$ direction to Newton's second law.

direction at velocity $-\mathbf{v}$ for an observer in S'). Figures 1-2 and 1-3 illustrate the conceptual differences between inertial and noninertial reference frames. Transformation of the position coordinates and the velocity components of S into those of S' is the *Galilean transformation*, Equations 1-2 and 1-3, respectively.

$$x' = x - vt \qquad y' = y \qquad z' = z \qquad t' = t \tag{1-2}$$

$$u'_x = u_x - v \qquad u'_y = u_y \qquad u'_z = u_z \tag{1-3}$$

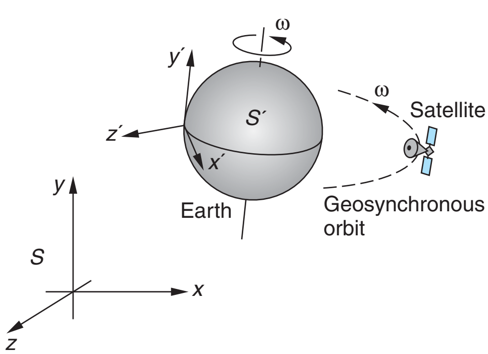


Figure 1-3 A geosynchronous satellite has an orbital angular velocity equal to that of Earth and, therefore, is always located above a particular point on Earth; i.e., it is at rest with respect to the surface of Earth. An observer in S accounts for the radial, or centripetal, acceleration $\mathbf{a}$ of the satellite as the result of the net force $\mathbf{F}_\mathbf{G}$. For an observer O′ at rest on Earth (in S'), however, $\mathbf{a}' = 0$ and $\mathbf{F}'_\mathbf{G} \neq m\mathbf{a}'$. To explain the acceleration being zero, observer O′ must add a pseudoforce $\mathbf{F}_\mathbf{p} = -\mathbf{F}_\mathbf{G}$.

Notice that differentiating Equation 1-3 yields the result $a' = a$ since $dv/dt = 0$ for constant v. Thus, $\mathbf{F} = m\mathbf{a} = \mathbf{F}'$. This is the invariance referred to above. Generalizing this result:

> Any reference frame that moves at constant velocity with respect to an inertial frame is also an inertial frame. Newton's laws of mechanics are invariant in all reference systems connected by a Galilean transformation.

Speed of Light

In about 1860 James Clerk Maxwell summarized the experimental observations of electricity and magnetism in a consistent set of four concise equations. Unlike Newton's laws of motion, Maxwell's equations are not invariant under a Galilean transformation between inertial reference frames (Figure 1-4). Since the Maxwell equations predict the existence of electromagnetic waves whose speed would be a particular value, $c = 1/\sqrt{\mu_0\epsilon_0} = 3.00 \times 10^8$ m/s, the excellent agreement between this number and the measured value of the speed of light[3] and between the predicted polarization properties of electromagnetic waves and those observed for light provided strong confirmation of the assumption that light was an electromagnetic wave and, therefore, traveled at speed c.[4]

That being the case, it was postulated in the nineteenth century that electromagnetic waves, like all other waves, propagated in a suitable material medium. The implication of this postulate was that the medium, called the *ether*, filled the entire universe, including the interior of matter. (The Greek philosopher Aristotle had first suggested that the universe was permeated with "ether" 2000 years earlier.) In this way the remarkable opportunity arose to establish experimentally the existence of the all-pervasive ether by measuring the speed of light c' relative to Earth as Earth moved relative to the ether at speed v, as would be predicted by Equation 1-3. The value of c was given by the Maxwell equations, and the speed of Earth relative to the ether, while not known, was assumed to be at least equal to its orbital speed around the Sun, about 30 km/s. Since the maximum observable effect is of the order v^2/c^2 and given this assumption $v^2/c^2 \approx 10^{-8}$, an experimental accuracy of about 1 part in 10^8 is necessary in order to detect Earth's motion relative to the ether. With a single exception, equipment and

Classical Concept Review

The concepts of classical relativity, frames of reference, and coordinate transformations—all important background to our discussions of special relativity—may not have been emphasized in many introductory courses. As an aid to a better understanding of the concepts of modern physics, we have included the *Classical Concept Review* on the book's Web site. As you proceed through *Modern Physics*, the icon in the margin will alert you to potentially helpful classical background pertinent to the adjacent topics.

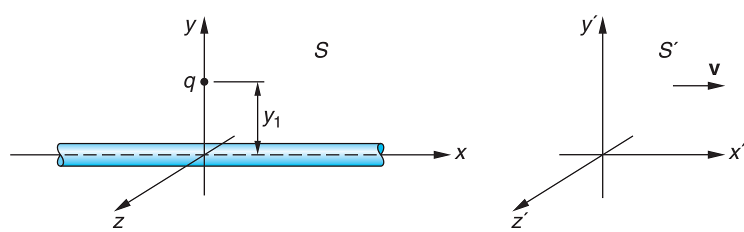


Figure 1-4 The observers in S and S' see identical electric fields $2k\lambda/y_1$ at a distance $y_1 = y_1'$ from an infinitely long wire carrying uniform charge λ per unit length. Observers in both S and S' measure a force $2kq\lambda/y_1$ on q due to the line of charge; however, the S' observer measures an additional force $-\mu_0\lambda v^2 q/(2\pi y_1)$ due to the magnetic field at y_1' arising from the motion of the wire in the $-x'$ direction. Thus, the electromagnetic force does not have the same form in different inertial systems, implying that Maxwell's equations are *not* invariant under a Galilean transformation.

techniques available at the time had an experimental accuracy of only about 1 part in 10^4, woefully insufficient to detect the predicted small effect. That single exception was the experiment of Michelson and Morley.[5]

Questions

1. What would the relative velocity of the inertial systems in Figure 1-4 need to be in order for the S' observer to measure no net electromagnetic force on the charge q?
2. Discuss why the very large value for the speed of the electromagnetic waves would imply that the ether be rigid, i.e., have a large bulk modulus.

The Michelson-Morley Experiment

All waves that were known to nineteenth-century scientists required a medium in order to propagate. Surface waves moving across the ocean obviously require the water. Similarly, waves move along a plucked guitar string, across the surface of a struck drumhead, through Earth after an earthquake, and, indeed, in all materials acted upon by suitable forces. The speed of the waves depends on the properties of the medium and is derived *relative to the medium*. For example, the speed of sound waves in air, i.e., their absolute motion relative to still air, can be measured. The Doppler effect for sound in air depends not only on the relative motion of the source and listener, but also on the motion of each relative to still air. Thus, it was natural for scientists of that time to expect the existence of some material like the ether to support the propagation of light and other electromagnetic waves *and* to expect that the absolute motion of Earth through the ether should be detectable, despite the fact that the ether had not been observed previously.

Albert A. Michelson, here playing pool in his later years, made the first accurate measurement of the speed of light while an instructor at the U.S. Naval Academy, where he had earlier been a cadet. [*AIP Emilio Segrè Visual Archives.*]

Michelson realized that although the effect of Earth's motion on the results of any "out-and–back" speed of light measurement, such as shown generically in Figure 1-5, would be too small to measure directly, it should be possible to measure v^2/c^2 by a difference measurement, using the interference property of the light waves as a sensitive "clock." The apparatus that he designed to make the measurement is called the *Michelson interferometer*. The purpose of the Michelson-Morley experiment was to measure the speed of light relative to the interferometer (i.e., relative to Earth), thereby detecting Earth's motion through the ether and thus verifying the latter's existence. To illustrate how the interferometer works and the reasoning behind the experiment, let us first describe an analogous situation set in more familiar surroundings.

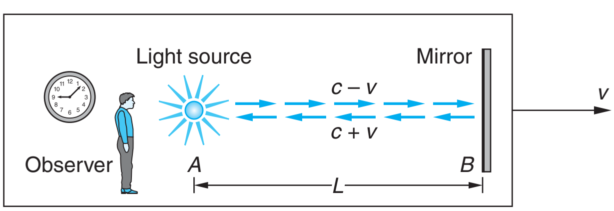


Figure 1-5 Light source, mirror, and observer are moving with speed v relative to the ether. According to classical theory, the speed of light c, relative to the ether, would be $c - v$ relative to the observer for light moving from the source toward the mirror and $c + v$ for light reflecting from the mirror back toward the source.

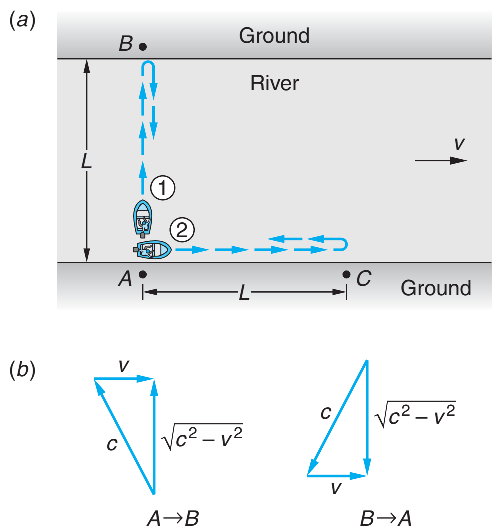


Figure 1-6 (*a*) The rowers both row at speed c in still water. (See Example 1-1.) The current in the river moves at speed v. Rower 1 goes from A to B and back to A, while rower 2 goes from A to C and back to A. (*b*) Rower 1 must point the bow upstream so that the sum of the velocity vectors $\mathbf{c} + \mathbf{v}$ results in the boat moving from A directly to B. His speed relative to the banks (i.e., points A and B) is then $(c^2 - v^2)^{1/2}$. The same is true on the return trip.

EXAMPLE 1-1 A Boat Race Two equally matched rowers race each other over courses as shown in Figure 1-6*a*. Each oarsman rows at speed c in still water; the current in the river moves at speed v. Boat 1 goes from A to B, a distance L, and back. Boat 2 goes from A to C, also a distance L, and back. A, B, and C are marks on the riverbank. Which boat wins the race, or is it a tie? (Assume $c > v$.)

SOLUTION

The winner is, of course, the boat that makes the round trip in the shortest time, so to discover which boat wins, we compute the time for each. Using the classical velocity transformation (Equations 1-3), the speed of 1 relative to the ground is $(c^2 - v^2)^{1/2}$, as shown in Figure 1-6*b*; thus the round-trip time t_1 for boat 1 is

$$\begin{aligned} t_1 &= t_{A\to B} + t_{B\to A} = \frac{L}{\sqrt{c^2 - v^2}} + \frac{L}{\sqrt{c^2 - v^2}} = \frac{2L}{\sqrt{c^2 - v^2}} \\ &= \frac{2L}{c\sqrt{1 - \dfrac{v^2}{c^2}}} = \frac{2L}{c}\left(1 - \frac{v^2}{c^2}\right)^{-1/2} \approx \frac{2L}{c}\left(1 + \frac{1}{2}\frac{v^2}{c^2} + \ldots\right) \end{aligned} \qquad \textbf{1-4}$$

where we have used the binomial expansion. Boat 2 moves downstream at speed $c + v$ relative to the ground and returns at $c - v$, also relative to the ground. The round-trip time t_2 is thus

$$\begin{aligned} t_2 &= \frac{L}{c + v} + \frac{L}{c - v} = \frac{2Lc}{c^2 - v^2} \\ &= \frac{2L}{c}\frac{1}{1 - \dfrac{v^2}{c^2}} \approx \frac{2L}{c}\left(1 + \frac{v^2}{c^2} + \ldots\right) \end{aligned} \qquad \textbf{1-5}$$

which, you may note, is the same result obtained in our discussion of the speed of light experiment in the Classical Concept Review.

The difference Δt between the round-trip times of the boats is then

$$\Delta t = t_2 - t_1 \approx \frac{2L}{c}\left(1 + \frac{v^2}{c^2}\right) - \frac{2L}{c}\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right) \approx \frac{Lv^2}{c^3} \qquad \textbf{1-6}$$

The quantity Lv^2/c^3 is always positive; therefore, $t_2 > t_1$ and rower 1 has the faster average speed and wins the race.

The Results Michelson and Morley carried out the experiment in 1887, repeating with a much-improved interferometer an inconclusive experiment that Michelson alone had performed in 1881 in Potsdam. The path length L on the new interferometer (Figure 1-7) was about 11 meters, obtained by a series of multiple reflections. Michelson's interferometer is shown schematically in Figure 1-8*a*. The field of view seen by the observer consists of parallel alternately bright and dark interference bands, called *fringes*, as illustrated in Figure 1-8*b*. The two light beams in the interferometer are exactly analogous to the two boats in Example 1-1, and Earth's motion through the ether was expected to introduce a time (phase) difference as given by

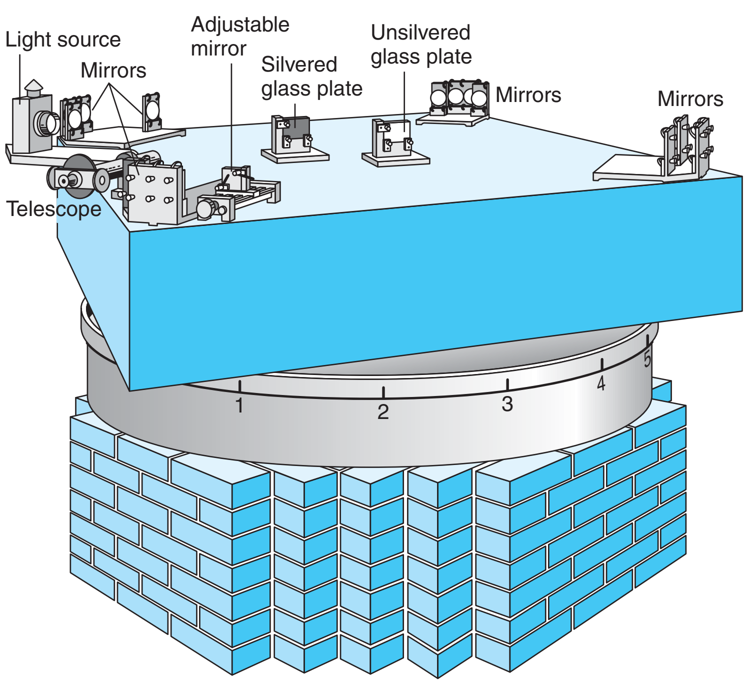


Figure 1-7 Drawing of Michelson-Morley apparatus used in their 1887 experiment. The optical parts were mounted on a 5 ft square sandstone slab, which was floated in mercury, thereby reducing the strains and vibrations during rotation that had affected the earlier experiments. Observations could be made in all directions by rotating the apparatus in the horizontal plane. [*From R. S. Shankland, "The Michelson-Morley Experiment," Copyright © November 1964 by Scientific American, Inc. All rights reserved.*]

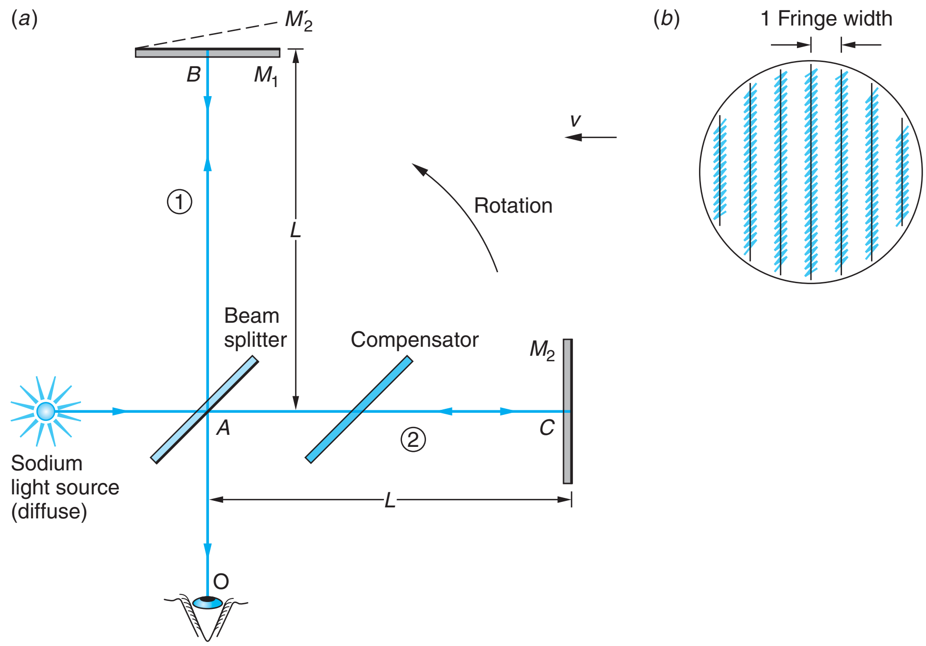


Figure 1-8 Michelson interferometer. (*a*) Yellow light from the sodium source is divided into two beams by the second surface of the partially reflective beam splitter at A, at which point the two beams are exactly in phase. The beams travel along the mutually perpendicular paths 1 and 2, reflect from mirrors M_1 and M_2, and return to A, where they recombine and are viewed by the observer. The compensator's purpose is to make the two paths of equal optical length, so that the lengths L contain the same number of light waves, by making both beams pass through two thicknesses of glass before recombining. M_2 is then tilted slightly so that it is not quite perpendicular to M_1. Thus, the observer O sees M_1 and M_2', the image of M_2 formed by the partially reflecting second surface of the beam splitter, forming a thin wedge-shaped film of air between them. The interference of the two recombining beams depends on the number of waves in each path, which in turn depends on (1) the length of each path and (2) the speed of light (relative to the instrument) in each path. Regardless of the value of that speed, the wedge-shaped air film between M_1 and M_2' results in an increasing path length for beam 2 relative to beam 1, looking from left to right across the observer's field of view; hence, the observer sees a series of parallel interference fringes as in (*b*), alternately yellow and black from constructive and destructive interference, respectively.

Equation 1-6. Rotating the interferometer through 90° doubles the time difference and changes the phase, causing the fringe pattern to shift by an amount ΔN. An improved system for rotating the apparatus was used in which the massive stone slab on which the interferometer was mounted floated on a pool of mercury. This dampened vibrations and enabled the experimenters to rotate the interferometer without introducing mechanical strains, both of which would cause changes in L and hence a shift in the fringes. Using a sodium light source with $\lambda = 590$ nm and assuming $v = 30$ km/s (i.e., Earth's orbital speed), ΔN was expected to be about 0.4 of the width of a fringe, about 40 times the minimum shift (0.01 fringe) that the interferometer was capable of detecting.

To Michelson's immense disappointment and that of most scientists of the time, the expected shift in the fringes did not occur. Instead, the shift observed was only about 0.01 fringe, i.e., approximately the experimental uncertainty of the apparatus. With characteristic reserve, Michelson described the results thus:[6]

> The actual displacement [of the fringes] was certainly less than the twentieth part [of 0.4 fringe], and probably less than the fortieth part. But since the displacement is proportional to the square of the velocity, the relative velocity of the earth and the ether is probably less than one-sixth the earth's orbital velocity and certainly less than one-fourth.

Michelson and Morley had placed an upper limit on Earth's motion relative to the ether of about 5 km/s. From this distance in time it is difficult for us to appreciate the devastating impact of this result. The then-accepted theory of light propagation could not be correct, and the ether as a favored frame of reference for Maxwell's equations was not tenable. The experiment was repeated by a number of people more than a dozen times under various conditions and with improved precision, and no shift has ever been found. In the most precise attempt, the upper limit on the relative velocity was lowered to 1.5 km/s by Georg Joos in 1930 using an interferometer with light paths much longer than Michelson's. Recent, high-precision variations of the experiment using laser beams have lowered the upper limit to 15 m/s.

Michelson interferometers with arms as long as 4 km are currently being used in the search for gravity waves. See Section 2-5.

More generally, on the basis of this and other experiments, we must conclude that Maxwell's equations are correct and that the speed of electromagnetic radiation is the same in all inertial reference systems independent of the motion of the source relative to the observer. This invariance of the speed of light between inertial reference frames means that there must be some relativity principle that applies to electromagnetism as well as to mechanics. That principle cannot be Newtonian relativity, which implies the dependence of the speed of light on the relative motion of the source and observer. It follows that the Galilean transformation of coordinates between inertial frames cannot be correct but must be replaced with a new coordinate transformation whose application preserves the invariance of the laws of electromagnetism. We then expect that the fundamental laws of mechanics, which were consistent with the old Galilean transformation, will require modification in order to be invariant under the new transformation. The theoretical derivation of that new transformation was a cornerstone of Einstein's development of special relativity.

More

A more complete description of the *Michelson-Morley experiment*, its interpretation, and the results of very recent versions can be found on the home page: www.whfreeman.com/tiplermodernphysics5e. See also Figures 1-9 through 1-11 here, as well as Equations 1-7 through 1-10.

1-2 Einstein's Postulates

In 1905, at the age of 26, Albert Einstein published several papers, among which was one on the electrodynamics of moving bodies.[11] In this paper, he postulated a more general principle of relativity that applied to the laws of both electrodynamics and mechanics. A consequence of this postulate is that absolute motion cannot be detected by any experiment. We can then consider the Michelson apparatus and Earth to be at rest. No fringe shift is expected when the interferometer is rotated 90°, since all directions are equivalent. The null result of the Michelson-Morley experiment is therefore to be expected. It should be pointed out that Einstein did not set out to explain the Michelson-Morley experiment. His theory arose from his considerations of the theory of electricity and magnetism and the unusual property of electromagnetic waves that they propagate in a vacuum. In his first paper, which contains the complete theory of special relativity, he made only a passing reference to the experimental attempts to detect Earth's motion through the ether, and in later years he could not recall whether he was aware of the details of the Michelson-Morley experiment before he published his theory.

The theory of special relativity was derived from two postulates proposed by Einstein in his 1905 paper:

Postulate 1. The laws of physics are the same in all inertial reference frames.

Postulate 2. The speed of light in a vacuum is equal to the value c, independent of the motion of the source.

Postulate 1 is an extension of the Newtonian principle of relativity to include all types of physical measurements (not just measurements in mechanics). It implies that no inertial system is preferred over any other; hence, absolute motion cannot be detected. Postulate 2 describes a common property of all waves. For example, the speed of sound waves does not depend on the motion of the sound source. When an approaching car sounds its horn, the frequency heard increases according to the Doppler effect, but the speed of the waves traveling through the air does not depend on the speed of the car. The speed of the waves depends only on the properties of the air, such as its temperature. The force of this postulate was to include light waves, for which experiments had found no propagation medium, together with all other waves, whose speed *was* known to be independent of the speed of the source. Recent analysis of the light curves of gamma-ray bursts that occur near the edge of the observable universe have shown the speed of light to be independent of the speed of the source to a precision of one part in 10^{20}.

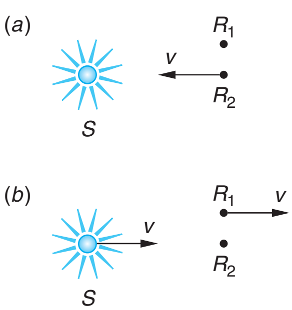


Figure 1-12 (*a*) Stationary light source S and a stationary observer R_1, with a second observer R_2 moving toward the source with speed v. (*b*) In the reference frame in which the observer R_2 is at rest, the light source S and observer R_1 move to the right with speed v. If absolute motion cannot be detected, the two views are equivalent. Since the speed of light does not depend on the motion of the source, observer R_2 measures the same value for that speed as observer R_1.

Although each postulate seems quite reasonable, many of the implications of the two together are surprising and seem to contradict common sense. One important implication of these postulates is that every observer measures the same value for the speed of light independent of the relative motion of the source and observer. Consider a light source S and two observers R_1, at rest relative to S, and R_2, moving toward S with speed v, as shown in Figure 1-12*a*. The speed of light measured by R_1 is $c = 3 \times 10^8$ m/s. What is the speed measured by R_2? The answer is *not* $c + v$, as one would expect based on Newtonian relativity. By postulate 1, Figure 1-12*a* is equivalent to Figure 1-12*b*, in which R_2 is at rest and the source S and R_1 are moving with speed v. That is, since absolute motion cannot be detected, it is not possible to say which is really moving and which is at rest. By postulate 2, the speed of light from a moving source is independent of the motion of the source. Thus, looking at Figure 1-12*b*, we see that R_2 measures the speed of light to be c, just as R_1 does. This result, that all observers measure the same value c for the speed of light, is often considered an alternative to Einstein's second postulate.

This result contradicts our intuition. Our intuitive ideas about relative velocities are approximations that hold only when the speeds are very small compared with the speed of light. Even in an airplane moving at the speed of sound, it is not possible to measure the speed of light accurately enough to distinguish the difference between the results c and $c + v$, where v is the speed of the plane. In order to make such a distinction, we must either move with a very great velocity (much greater than that of sound) or make extremely accurate measurements, as in the Michelson-Morley experiment, and when we do, we will find, as Einstein pointed out in his original relativity paper, that the contradictions are "only apparently irreconcilable."

Events and Observers

In considering the consequences of Einstein's postulates in greater depth, i.e., in developing the theory of special relativity, we need to be certain that meanings of some important terms are crystal clear. First, there is the concept of an *event*. A physical event is something that happens, like the closing of a door, a lightning strike, the collision of two particles, your birth, or the explosion of a star. Every event occurs at some point in space and at some instant in time, but it is very important to recognize that events are independent of the particular inertial reference frame that we might use to describe them. Events do not "belong" to any reference frame.

Events are described by *observers* who do belong to particular inertial frames of reference. Observers could be people (as in Section 1-1), electronic instruments, or other suitable recorders, but for our discussions in special relativity we are going to be very specific. Strictly speaking, the observer will be an array of recording clocks located throughout the inertial reference system. It may be helpful for you to think of the observer as a person who goes around reading out the memories of the recording clocks or receives records that have been transmitted from distant clocks, but always keep in mind that in reporting events, such a person is strictly limited to summarizing the data collected from the clock memories. The travel time of light precludes him from including in his report distant events that he may have seen by eye! It is in this sense that we will be using the word *observer* in our discussions.

Each inertial reference frame may be thought of as being formed by a cubic three-dimensional lattice made of identical measuring rods (e.g., meter sticks) with a recording clock at each intersection as illustrated in Figure 1-13. The clocks are all identical, and we, of course, want them all to read the "same time" as one another at any instant; i.e., they must be *synchronized*. There are many ways to accomplish synchronization of the clocks, but a very straightforward way, made possible by the second postulate, is to use one of the clocks in the lattice as a standard, or *reference clock*. For convenience we will also use the location of the reference clock in the lattice as the coordinate origin for the reference frame. The reference clock is started with its indicator (hands, pointer, digital display) set at zero. At the instant it starts, it also sends out a flash of light that spreads out as a spherical wave in all directions. When the flash from the reference clock reaches the lattice clocks 1 meter away (notice that in Figure 1-13 there are six of them, two of which are off the edges of the figure), we want their indicators to read the time required for light to travel 1 m (= 1/299,792,458 s). This can be done simply by having an observer at each clock set that time on the indicator and then having the flash from the reference clock start them as it passes. The clocks 1 m from the origin now display the same time as the reference clock; i.e., they are all synchronized. In a similar fashion, all of the clocks throughout the inertial frame can be synchronized since the distance of any clock from the reference clock can be calculated from the space coordinates of its position in the lattice and the initial setting of its indicator will be the corresponding travel time for the reference light flash. This procedure can be used to synchronize the clocks in any inertial frame, *but* it does not synchronize the clocks in reference frames that move with respect to one another. Indeed, as we shall see shortly, clocks in relatively moving frames cannot in general be synchronized with one another.

(*Top*) Albert Einstein in 1905 at the Bern, Switzerland, patent office. [*Hebrew University of Jerusalem Albert Einstein Archives, courtesy AIP Emilio Segrè Visual Archives.*] (*Bottom*) Clock tower and electric trolley in Bern on Kramstrasse, the street on which Einstein lived. If you are on the trolley moving away from the clock and look back at it, the light you see must catch up with you. If you move at nearly the speed of light, the clock you see will be slow. In this, Einstein saw a clue to the variability of time itself. [*Underwood & Underwood/CORBIS.*]

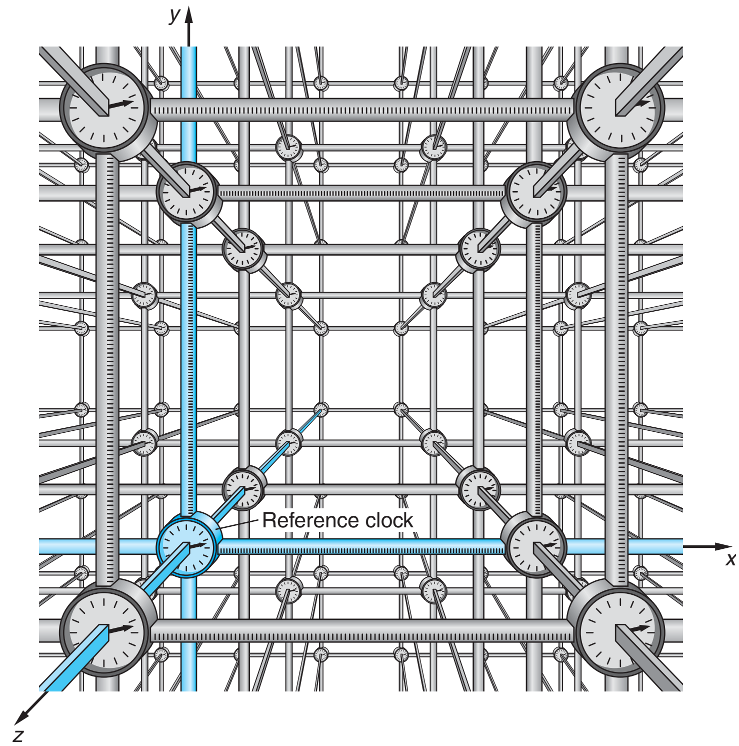


Figure 1-13 Inertial reference frame formed from a lattice of measuring rods with a clock at each intersection. The clocks are all synchronized using a reference clock. In this diagram the measuring rods are shown to be 1 m long, but they could all be 1 cm, 1 μm, or 1 km as required by the scale and precision of the measurements being considered. The three space dimensions are the clock positions. The fourth spacetime dimension, time, is shown by indicator readings on the clocks.

When an event occurs, its location and time are recorded instantly by the nearest clock. Suppose that an atom located at $x = 2$ m, $y = 3$ m, $z = 4$ m in Figure 1-13 emits a tiny flash of light at $t = 21$ s on the clock at that location. That event is recorded in space and in time or, as we will henceforth refer to it, in the *spacetime* coordinate system with the numbers (2,3,4,21). The observer may read out and analyze these data at his leisure, within the limits set by the information transmission time (i.e., the light travel time) from distant clocks. For example, the path of a particle moving through the lattice is revealed by analysis of the records showing the particle's time of passage at each clock's location. Distances between successive locations and the corresponding time differences make possible the determination of the particle's velocity. Similar records of the spacetime coordinates of the particle's path can, of course, also be made in any inertial frame moving relative to ours, but to compare the distances and time intervals measured in the two frames requires that we consider carefully the relativity of simultaneity.

Relativity of Simultaneity

Einstein's postulates lead to a number of predictions about measurements made by observers in inertial frames moving relative to one another that initially seem very strange, including some that appear paradoxical. Even so, these predictions have been experimentally verified; and nearly without exception, every paradox is resolved by an understanding of the *relativity of simultaneity*, which states that

> Two spatially separated events simultaneous in one reference frame are not, in general, simultaneous in another inertial frame moving relative to the first.

A corollary to this is that

> Clocks synchronized in one reference frame are not, in general, synchronized in another inertial frame moving relative to the first.

What do we mean by simultaneous events? Suppose two observers, both in the inertial frame S at different locations A and B, agree to explode bombs at time t_o (remember, we have synchronized all of the clocks in S). The clock at C, equidistant from A and B, will record the arrival of light from the explosions at the same instant, i.e., simultaneously. Other clocks in S will record the arrival of light from A or B first, depending on their locations, but after correcting for the time the light takes to reach each clock, the data recorded by each would lead an observer to conclude that the explosions were simultaneous. *We will thus define two events to be simultaneous in an inertial reference frame if the light signals from the events reach an observer halfway between them at the same time as recorded by a clock at that location*, called a local clock.

Einstein's Example To show that two events that are simultaneous in frame S are not simultaneous in another frame S' moving relative to S, we use an example introduced by Einstein. A train is moving with speed v past a station platform. We have observers located at A', B', and C' at the front, back, and middle of the train. (We consider the train to be at rest in S' and the platform in S.) We now suppose that the train and platform are struck by lightning at the front and back of the train and that the lightning bolts are simultaneous in the frame of the platform (S; Figure 1-14*a*). That is, an observer located at C halfway between positions A and B, where lightning strikes, observes the two flashes at the same time. It is convenient to suppose that the

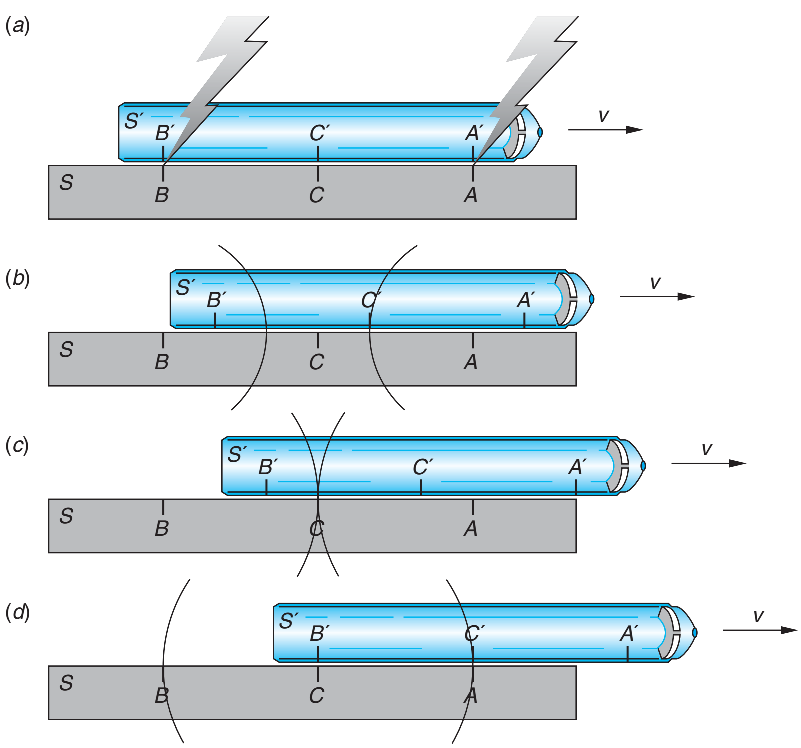


Figure 1-14 Lightning bolts strike the front and rear of the train, scorching both the train and the platform, as the train (frame S') moves past the platform (system S) at speed v. (*a*) The strikes are simultaneous in S, reaching the C observer located midway between the events at the same instant as recorded by the clock at C as shown in (*c*). In S' the flash from the front of the train is recorded by the C' clock, located midway between the scorch marks on the train, before that from the rear of the train (*b* and *d*, respectively). Thus, the C' observer concludes that the strikes were not simultaneous.

lightning scorches both the train and the platform so that the events can be easily located in each reference frame. Since C' is in the middle of the train, halfway between the places on the train that are scorched, the events are simultaneous in S' only if the clock at C' records the flashes at the same time. However, the clock at C' records the flash from the front of the train before the flash from the back. In frame S, when the light from the front flash reaches the observer at C', the train has moved some distance toward A, so that the flash from the back has not yet reached C', as indicated in Figure 1-14*b*. The observer at C' must therefore conclude that the events are not simultaneous, but that the front of the train was struck before the back. Figures 1-14*c* and 1-14*d* illustrate, respectively, the subsequent simultaneous arrival of the flashes at C and the still-later arrival of the flash from the rear of the train at C'. As we have discussed, all observers in S' on the train will agree with the observer C' when they have corrected for the time it takes light to reach them.

The corollary can also be demonstrated with a similar example. Again consider the train to be at rest in S' that moves past the platform, at rest in S, with speed v. Figure 1-15 shows three of the clocks in the S lattice and three of those in the S' lattice. The clocks in each system's lattice have been synchronized in the manner that was described earlier, but those in S are not synchronized with those in S'. The observer at C midway between A and B on the platform announces that light sources at A and B will flash when the clocks at those locations read t_o (Figure 1-15*a*). The observer at C', positioned midway between A' and B', notes the arrival of the light flash from the front of the train (Figure 1-15*b*) *before* the arrival of the one from the rear (Figure 1-15*d*). Observer C' thus concludes that if the flashes were each emitted at t_o on the local clocks, as announced, then the clocks at A and B are not synchronized. All observers in S' would agree with that conclusion after correcting for the time of light travel. The clock located at C records the arrival of the two flashes simultaneously, of course, since the clocks in S are synchronized (Figure 1-15*c*).

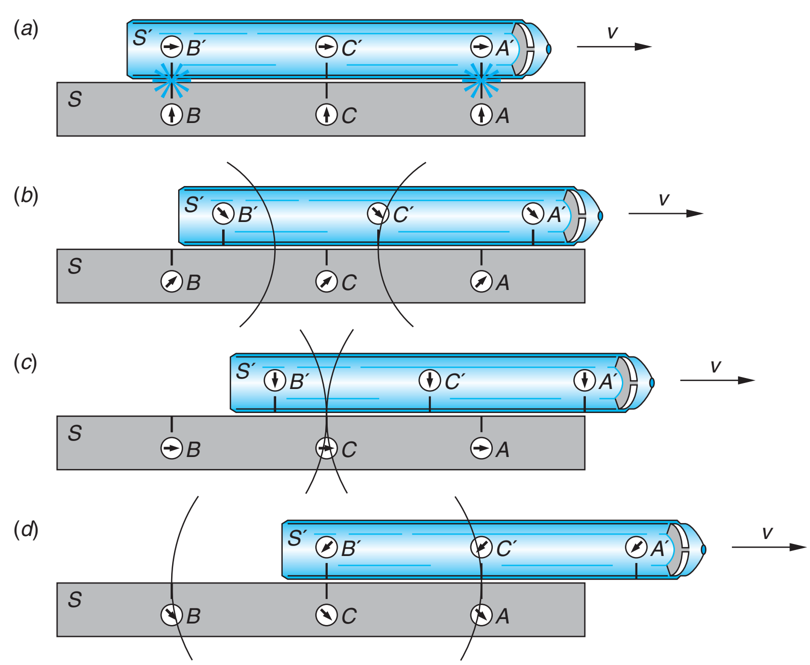


Figure 1-15 (*a*) Light flashes originate simultaneously at clocks A and B, synchronized in S. (*b*) The clock at C', midway between A' and B' on the moving train, records the arrival of the flash from A before the flash from B shown in (*d*). Since the observer in S announced that the flashes were triggered at t_o on the local clocks, the observer at C' concludes that the local clocks at A and B did not read t_o simultaneously; i.e., they were not synchronized. The simultaneous arrival of the flashes at C is shown in (*c*).

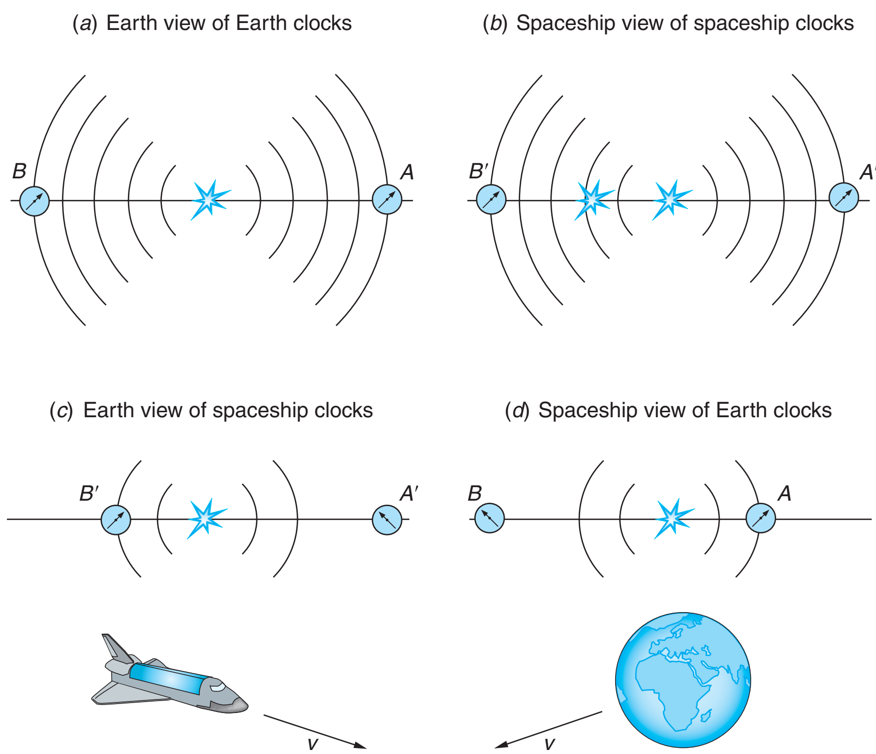


Figure 1-16 A light flash occurs on Earth midway between two Earth clocks. At the instant of the flash the midpoint of a passing spaceship coincides with the light source. (*a*) The Earth clocks record the lights' arrival simultaneously and are thus synchronized. (*b*) Clocks at both ends of the spaceship also record the lights' arrival simultaneously (Einstein's second postulate) and they, too, are synchronized. (*c*) However, the Earth observer sees the light reach the clock at B' *before* the light reaches the clock at A'. Since the spaceship clocks read the same time when the light arrives, the Earth observer concludes that the clocks at A' and B' are not synchronized. (*d*) The spaceship observer similarly concludes that the Earth clocks are not synchronized.

Notice, too, in Figure 1-15 that C' also concludes that the clock at A is ahead of the clock at B. This is important, and we will return to it in more detail in the next section. Figure 1-16 illustrates the relativity of simultaneity from a different perspective.

Questions

3. In addition to that described above, what would be another possible method of synchronizing all of the clocks in an inertial reference system?
4. Using Figure 1-16*d*, explain how the spaceship observer concludes that Earth clocks are not synchronized.

1-3 The Lorentz Transformation

We now consider a very important consequence of Einstein's postulates, the general relation between the spacetime coordinates x, y, z, and t of an event as seen in reference frame S and the coordinates x', y', z', and t' of the same event as seen in reference frame S', which is moving with uniform velocity relative to S. For simplicity we will

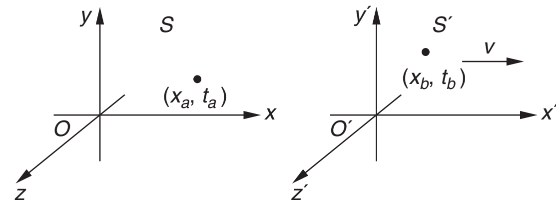


Figure 1-17 Two inertial frames S and S' with the latter moving at speed v in $+x$ direction of system S. Each set of axes shown is simply the coordinate axes of a lattice like that in Figure 1-13. Remember, there is a clock at each intersection. A short time before, the times represented by this diagram O and O' were coincident and the lattices of S and S' were intermeshed.

consider only the special case in which the origins of the two coordinate systems are coincident at time $t = t' = 0$ and S' is moving, relative to S, with speed v along the x (or x') axis and with the y' and z' axes parallel, respectively, to the y and z axes, as shown in Figure 1-17. As we discussed earlier (Equation 1-2), the classical Galilean coordinate transformation is

$$x' = x - vt \qquad y' = y \qquad z' = z \qquad t' = t \tag{1-2}$$

which expresses coordinate measurements made by an observer in S' in terms of those measured by an observer in S. The inverse transformation is

$$x = x' + vt' \qquad y = y' \qquad z = z' \qquad t = t'$$

and simply reflects the fact that the sign of the relative velocity of the reference frames is different for the two observers. The corresponding classical velocity transformation was given in Equation 1-3 and the acceleration, as we saw earlier, is invariant under a Galilean transformation. (For the rest of the discussion we will ignore the equations for y and z, which do not change in this special case of motion along the x and x' axes.) These equations are consistent with experiment as long as v is much less than c.

It should be clear that the classical velocity transformation is not consistent with the Einstein postulates of special relativity. If light moves along the x axis with speed c in S, Equation 1-3 implies that the speed in S' is $u'_x = c - v$ rather than $u'_x = c$. The Galilean transformation equations must therefore be modified to be consistent with Einstein's postulates, but the result must reduce to the classical equations when v is much less than c. We will give a brief outline of one method of obtaining the relativistic transformation that is called the *Lorentz transformation*, so named because of its original discovery by H. A. Lorentz.[12] We assume the equation for x' to be of the form

$$x' = \gamma(x - vt) \tag{1-11}$$

where γ is a constant that can depend upon v and c but not on the coordinates. If this equation is to reduce to the classical one, γ must approach 1 as v/c approaches 0. The inverse transformation must look the same except for the sign of the velocity:

$$x = \gamma(x' + vt') \tag{1-12}$$

With the arrangement of the axes in Figure 1-17, there is no relative motion of the frames in the y and z directions; hence $y' = y$ and $z' = z$. However, insertion of the as yet unknown multiplier γ modifies the classical transformation of time, $t' = t$. To see this, we substitute x' from Equation 1-11 into Equation 1-12 and solve for t'. The result is

$$t' = \gamma\left[t + \frac{(1-\gamma^2)}{\gamma^2}\frac{x}{v}\right] \tag{1-13}$$

Now let a flash of light start from the origin of S at $t = 0$. Since we have assumed that the origins coincide at $t = t' = 0$, the flash also starts at the origin of S' at $t' = 0$. The flash expands from *both* origins as a spherical wave. The equation for the wave front according to an observer in S is

$$x^2 + y^2 + z^2 = c^2t^2 \tag{1-14}$$

and according to an observer in S', it is

$$x'^2 + y'^2 + z'^2 = c^2t'^2 \tag{1-15}$$

where both equations are consistent with the second postulate. Consistency with the first postulate means that the relativistic transformation that we seek must transform Equation 1-14 into Equation 1-15 and vice versa. For example, substituting Equations 1-11 and 1-13 into 1-15 results in Equation 1-14 *if*

$$\gamma = \frac{1}{\sqrt{1-\dfrac{v^2}{c^2}}} = \frac{1}{\sqrt{1-\beta^2}} \tag{1-16}$$

where $\beta = v/c$. Notice that $\gamma = 1$ for $v = 0$ and $\gamma \to \infty$ for $v = c$. How this is done is illustrated in Example 1-2 below.

EXAMPLE 1-2 Relativistic Transformation Multiplier γ Show that γ must be given by Equation 1-16 if Equation 1-15 is to be transformed into Equation 1-14 consistent with Einstein's first postulate.

SOLUTION

Substituting Equations 1-11 and 1-13 into Equation 1-15 and noting that $y' = y$ and $z' = z$ in this case yields

$$\gamma^2(x - vt)^2 + y^2 + z^2 = c^2\gamma^2\left[t + \frac{1-\gamma^2}{\gamma^2}\frac{x}{v}\right]^2 \tag{1-17}$$

To be consistent with the first postulate, Equation 1-15 must be identical to 1-12. This requires that the coefficient of the x^2 term in Equation 1-17 be equal to 1, that of the t^2 term be equal to c^2, and that of the xt term be equal to 0. Any of those conditions can be used to determine γ, and all yield the same result. Using, for example, the coefficient of x^2, we have from Equation 1-17 that

$$\gamma^2 - c^2\gamma^2\frac{(1-\gamma^2)^2}{\gamma^4v^2} = 1$$

which can be rearranged to

$$-c^2\frac{(1-\gamma^2)^2}{\gamma^2 v^2} = (1-\gamma^2)$$

Canceling $1-\gamma^2$ on both sides and solving for γ yields

$$\gamma = \frac{1}{\sqrt{1-\dfrac{v^2}{c^2}}}$$

With the value for γ found in Example 1-2, Equation 1-13 can be written in a somewhat simpler form, and with it the complete Lorentz transformation becomes

$$x' = \gamma(x - vt) \qquad y' = y$$
$$t' = \gamma\left(t - \frac{vx}{c^2}\right) \qquad z' = z \tag{1-18}$$

and the inverse

$$x = \gamma(x' + vt') \qquad y = y'$$
$$t = \gamma\left(t' + \frac{vx'}{c^2}\right) \qquad z = z' \tag{1-19}$$

with

$$\gamma = \frac{1}{\sqrt{1-\beta^2}}$$

EXAMPLE 1-3 Transformation of Time Intervals The arrivals of two cosmic-ray μ leptons (muons) are recorded by detectors in the laboratory, one at time t_a at location x_a and the second at time t_b at location x_b in the laboratory reference frame, S in Figure 1-17. What is the time interval between those two events in system S', which moves relative to S at speed v?

SOLUTION

Applying the time coordinate transformation from Equation 1-18,

$$t'_b - t'_a = \gamma\left(t_b - \frac{vx_b}{c^2}\right) - \gamma\left(t_a - \frac{vx_a}{c^2}\right)$$
$$t'_b - t'_a = \gamma(t_b - t_a) - \frac{\gamma v}{c^2}(x_b - x_a) \tag{1-20}$$

We see that the time interval measured in S' depends not just on the corresponding time interval in S, but also on the spatial separation of the clocks in S that measured the interval. This result should not come as a total surprise, since we have already discovered that although the clocks in S are synchronized with each other, they are not, in general, synchronized for observers in other inertial frames.

Special Case 1

If the two events happen to occur at the same location in S, i.e., $x_a = x_b$, then $(t_b - t_a)$, the time interval measured on a clock located at the events, is called the *proper time interval*. Notice that since $\gamma > 1$ for all frames moving relative to S, the proper time interval is the *minimum* time interval that can be measured between those events.

Special Case 2

Does an inertial frame exist for which the events described above would be measured as being simultaneous? Since the question has been asked, you probably suspect that the answer is yes, and you are right. The two events will be simultaneous in a system S'' for which $t''_b - t''_a = 0$, i.e., when

$$\gamma(t_b - t_a) = \frac{\gamma v}{c^2}(x_b - x_a)$$

or when

$$\beta = \frac{v}{c} = \left(\frac{t_b - t_a}{x_b - x_a}\right)c \tag{1-21}$$

Notice that $(x_b - x_a)/c$ = time for a light beam to travel from x_a to x_b; thus we can characterize S'' as being that system whose speed relative to S is that fraction of c given by the time interval between the events divided by the travel time of light between them. (Note, too, that $c(t_b - t_a) > (x_b - x_a)$ implies that $\beta > 1$, a nonphysical situation that we will discuss in Section 1-4.)

While it is possible for us to get along in special relativity without the Lorentz transformation, it has an application that is quite valuable: it enables the spacetime coordinates of events measured by the measuring rods and clocks in the reference frame of one observer to be translated into the corresponding coordinates determined by the measuring rods and clocks of an observer in another inertial frame. As we will see in Section 1-4, such transformations lead to some startling results.

Relativistic Velocity Transformations

The transformation for velocities in special relativity can be obtained by differentiation of the Lorentz transformation, keeping in mind the definition of the velocity. Suppose a particle moves in S with velocity $\mathbf{u}$ whose components are $u_x = dx/dt$, $u_y = dy/dt$, and $u_z = dz/dt$. An observer in S' would measure the components $u'_x = dx'/dt'$, $u'_y = dy'/dt'$, and $u'_z = dz'/dt'$. Using the transformation equations, we obtain

$$dx' = \gamma(dx - v\,dt) \qquad dy' = dy$$

$$dt' = \gamma\left(dt - \frac{v\,dx}{c^2}\right) \qquad dz' = dz$$

from which we see that u'_x is given by

$$u'_x = \frac{dx'}{dt'} = \frac{\gamma(dx - vdt)}{\gamma\left(dt - \frac{vdx}{c^2}\right)} = \frac{(dx/dt - v)}{1 - \frac{v}{c^2}\frac{dx}{dt}}$$

or

$$u'_x = \frac{u_x - v}{1 - \frac{vu_x}{c^2}} \tag{1-22}$$

and, if a particle has velocity components in the y and z directions, it is not difficult to find the components in S' in a similar manner.

$$u'_y = \frac{u_y}{\gamma\left(1 - \frac{vu_x}{c^2}\right)} \qquad u'_z = \frac{u_z}{\gamma\left(1 - \frac{vu_z}{c^2}\right)}$$

Remember that this form of the velocity transformation is specific to the arrangement of the coordinate axes in Figure 1-17. Note, too, that when $v \ll c$, i.e., when $\beta = v/c \approx 0$, the relativistic velocity transforms reduce to the classical velocity addition of Equation 1-3. Likewise, the inverse velocity transformation is

$$u_x = \frac{u'_x + v}{\left(1 + \frac{vu'_x}{c^2}\right)} \qquad u_y = \frac{u'_y}{\gamma\left(1 + \frac{vu'_x}{c^2}\right)} \qquad u_z = \frac{u'_z}{\gamma\left(1 + \frac{vu'_x}{c^2}\right)} \tag{1-23}$$

EXAMPLE 1-4 Relative Speeds of Cosmic Rays Suppose that two cosmic ray protons approach Earth from opposite directions, as shown in Figure 1-18*a*. The speeds relative to Earth are measured to be $v_1 = 0.6c$ and $v_2 = -0.8c$. What is Earth's velocity relative to each proton, and what is the velocity of each proton relative to the other?

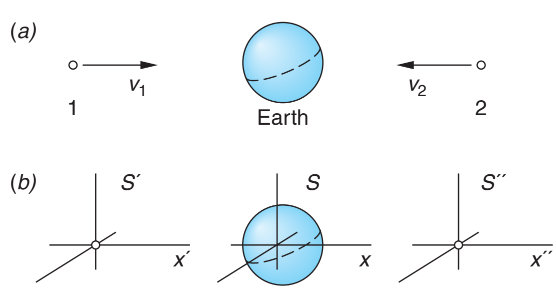


Figure 1-18 (*a*) Two cosmic ray protons approach Earth from opposite directions at speeds v_1 and v_2 with respect to Earth. (*b*) Attaching an inertial frame to each particle and Earth enables one to visualize the several relative speeds involved and apply the velocity transformation correctly.

SOLUTION

Consider each particle and Earth to be inertial reference frames S', S'', and S with their respective x axes parallel as in Figure 1-18*b*. With this arrangement $v_1 = u_{1x} = 0.6c$ and $v_2 = u_{2x} = -0.8c$. Thus, the speed of Earth measured in S' is $v'_{Ex} = -0.6c$ and the speed of Earth measured in S'' is $v''_{Ex} = 0.8c$.

To find the speed of proton 2 with respect to proton 1, we apply Equation 1-22 to compute u'_{2x}, i.e., the speed of particle 2 in S'. Its speed in S has been measured to be $u_{2x} = -0.8c$, where the S' system has relative speed $v_1 = 0.6c$ with respect to S. Thus, substituting into Equation 1-22, we obtain

$$u'_{2x} = \frac{-0.8c - (0.6c)}{1 - (0.6c)(-0.8c)/c^2} = \frac{-1.4c}{1.48} = -0.95c$$

and the first proton measures the second to be approaching (moving in the $-x'$ direction) at $0.95c$.

The observer in S'' must of course make a consistent measurement, i.e., find the speed of proton 1 to be $0.95c$ in the $+x''$ direction. This can be readily shown by a second application of Equation 1-22 to compute u''_{1x}:

$$u''_{1x} = \frac{0.6c - (-0.8c)}{1 - (0.6c)(-0.8c)/c^2} = \frac{1.4c}{1.48} = 0.95c$$

Questions

5. The Lorentz transformation for y and z is the same as the classical result: $y = y'$ and $z = z'$. Yet the relativistic velocity transformation does not give the classical result $u_y = u'_y$ and $u_z = u'_z$. Explain.
6. Since the velocity components of a moving particle are different in relatively moving frames, the *directions* of the velocity vectors are also different in general. Explain why the fact that observers in S and S' measure different directions for a particle's motion is not an inconsistency in their observations.

Spacetime Diagrams

The relativistic discovery that time intervals between events are not the same for all observers in different inertial reference frames underscores the four-dimensional character of spacetime. With the diagrams that we have used thus far, it is difficult to depict and visualize on the two-dimensional page events that occur at different times, since each diagram is equivalent to a snapshot of spacetime at a particular instant. Showing events as a function of time typically requires a series of diagrams, such as Figures 1-14, 1-15, and 1-16, but even then our attention tends to be drawn to the space coordinate systems rather than the events, whereas it is the *events* that are fundamental. This difficulty is removed in special relativity with a simple yet powerful graphing method called the *spacetime diagram*. (This is just a new name given to the t vs. x graphs that you first began to use when you discussed motion in introductory physics.) On the spacetime diagram we can graph both the space and time coordinates of many events in one or more inertial frames, albeit with one limitation. Since the page offers only two dimensions for graphing, we suppress, or ignore for now, two of the space dimensions, in particular y and z. With our choice of the relative motion of inertial frames along the x axis, $y' = y$ and $z' = z$ anyhow. (This is one of the reasons we made that convenient choice a few pages back, the other reason being mathematical simplicity.) This means that for the time being, we are limiting our attention to one space dimension and to time, i.e., to events that occur, regardless of

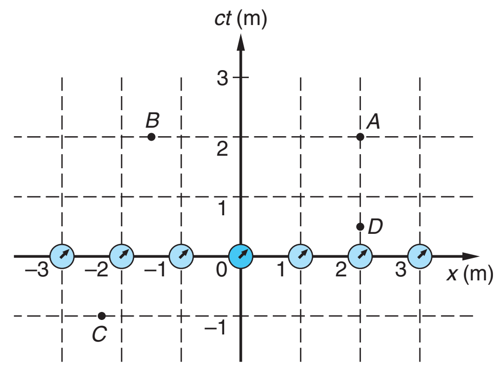


Figure 1-19 Spacetime diagram for an inertial reference frame S. Two of the space dimensions (y and z) are suppressed. The units on both the space and time axes are the same, meters. A meter of time means the time required for light to travel one meter, i.e., 3.3×10^{-9} s.

when, along one line in space. Should we need the other two dimensions, e.g., in a consideration of velocity vector transformations, we can always use the Lorentz transformation equations.

In a spacetime diagram the space location of each event is plotted along the x axis horizontally and the time is plotted vertically. From the three-dimensional array of measuring rods and clocks in Figure 1-13, we will use only those located on the x axis, as in Figure 1-19. (See, things are simpler already!) Since events that exhibit relativistic effects generally occur at high speeds, it will be convenient to multiply the time scale by the speed of light (a constant), which enables us to use the same units and scale on both the space and time axes, e.g., meters of distance and meters of light travel time.[13] The time axis is, therefore, c times the time t in seconds, i.e., ct. As we will see shortly, this choice prevents events from clustering about the axes and makes possible the straightforward addition of other inertial frames into the diagram.

As time advances, notice that in Figure 1-19 each clock in the array moves vertically upward along the dotted lines. Thus, as events A, B, C, and D occur in spacetime, one of the clocks of the array is at (or very near) each event when it happens. Remember that the clocks in the reference frame are synchronized, and so the difference in the readings of clocks located at each event records the *proper time interval* between the events. (See Example 1-3.) In the figure, events A and D occur at the same place ($x = 2$ m) but at different times. The time interval between them measured on clock 2 is the proper time interval since clock 2 is located at *both* events. Events A and B occur at different locations but at the same time (i.e., simultaneously in this frame). Event C occurred before the present since $ct = -1$ m. For this discussion we will consider the time that the coordinate origins coincide, $ct = ct' = 0$, to be the present.

Worldlines in Spacetime Particles moving in space trace out a line in the spacetime diagram called the *worldline* of the particle. The worldline is the "trajectory" of the particle on a ct versus x graph. To illustrate, consider four particles moving in space (not spacetime) as shown in Figure 1-20a, which shows the array of synchronized clocks on the x axis and the space trajectories of four particles, each starting at $x = 0$ and moving at some constant speed during 3 m of time. Figure 1-20b shows the worldline for each of the particles in spacetime. Notice that constant speed means that the worldline has constant slope; i.e., it is a straight line (slope $= \Delta t/\Delta x = 1/(\Delta x/\Delta t) = 1/\text{speed}$).

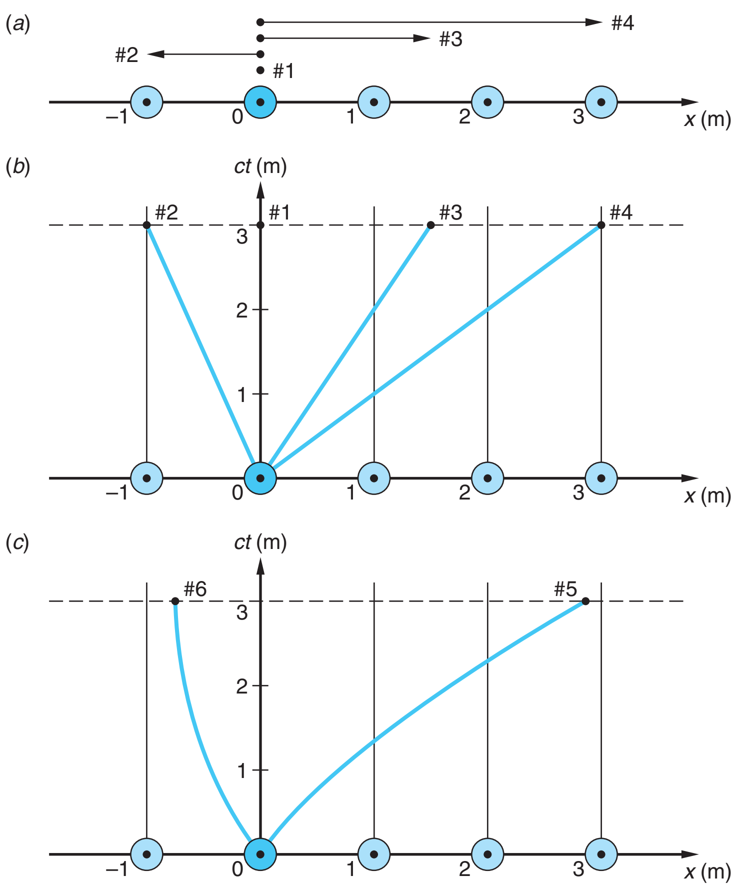


Figure 1-20 (*a*) The space trajectories of four particles with various constant speeds. Note that particle 1 has a speed of zero and particle 2 moves in the $-x$ direction. The worldlines of the particles are straight lines. (*b*) The worldline of particle 1 is also the *ct* axis since that particle remains at $x = 0$. The constant slopes are a consequence of the constant speeds. (*c*) For accelerating particles 5 and 6 [not shown in (*a*)], the worldlines are curved, the slope at any point yielding the instantaneous speed.

That was also the case when you first encountered elapsed time versus displacement graphs in introductory physics. Even then, you were plotting spacetime graphs and drawing worldlines! If the particle is accelerating—either speeding up like particle 5 in Figure 1-20*c* or slowing down like particle 6—the worldlines are curved. Thus, the worldline is the record of the particle's travel through spacetime, giving its speed (= 1/slope) and acceleration (= 1/rate at which the slope changes) at every instant.

EXAMPLE 1-5 Computing Speeds in Spacetime Find the speed u of particle 3 in Figure 1-20.

SOLUTION

The speed $u = \Delta x/\Delta t = 1/\text{slope}$, where we have $\Delta x = 1.5 - 0 = 1.5$ m and $\Delta ct = c \cdot \Delta t = 3.0 - 0 = 3.0$ m (from Figure 1-20). Thus, $\Delta t = (3.0/c) = (3.0/3.0 \times 10^8) = 10^{-8}$ s and $u = 1.5\text{ m}/10^{-8}\text{ s} = 0.5c$.

The speed of particle 4, computed as shown in Example 1-5, turns out to be c, the speed of light. (Particle 4 is a light pulse.) The slope of its worldline $\Delta(ct)/\Delta x = 3\text{ m}/3\text{ m} = 1$. Similarly, the slope of the worldline of a light pulse moving in the $-x$ direction is -1. Since relativity limits the speed of particles with mass to less than c,

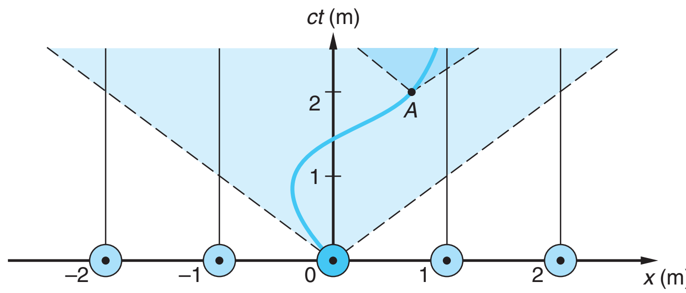


Figure 1-21 The speed-of-light limit to the speeds of particles limits the slopes of worldlines for particles that move through $x = 0$ at $ct = 0$ to the shaded area of spacetime, i.e., to slopes <-1 and $>+1$. The dashed lines are worldlines of light flashes moving in the $-x$ and $+x$ directions. The curved worldline of the particle shown has the same limits at every instant. Notice that the particle's speed = 1/slope.

as we will see in Chapter 2, the slopes of worldlines for particles that move through $x = 0$ at $ct = 0$ are limited to the larger shaded triangle in Figure 1-21. The same limits to the slope apply at every point along a particle's worldline, such as point A on the curved spacetime trajectory in Figure 1-21. This means that the particle's possible worldlines for times greater than $ct = 2$ m must lie within the heavily shaded triangle.

Analyzing events and motion in inertial systems that are in relative motion can now be accomplished more easily than with diagrams such as Figures 1-14 through 1-18. Suppose we have two inertial frames S and S' with S' moving in the $+x$ direction of S at speed v as in those figures. The clocks in both systems are started at $t = t' = 0$ (the present) as the two origins $x = 0$ and $x' = 0$ coincide, and, as before, observers in each system have synchronized the clocks in their respective systems. The spacetime diagram for S is, of course, like that in Figure 1-19, but how does S' appear in that diagram, i.e., with respect to an observer in S? Consider that as the origin of S' (i.e., the point where $x' = 0$) moves in S, its worldline is the ct' axis since the ct' axis is the locus of all points with $x' = 0$ (just as the ct axis is the locus of points with $x = 0$.) Thus, the slope of the ct' axis as seen by an observer in S can be found from Equation 1-18, the Lorentz transformation, as follows:

$$x' = \gamma(x - vt) = 0 \qquad \text{for} \qquad x' = 0$$

or

$$x = vt = (v/c)(ct) = \beta ct$$

and

$$ct = (1/\beta)x$$

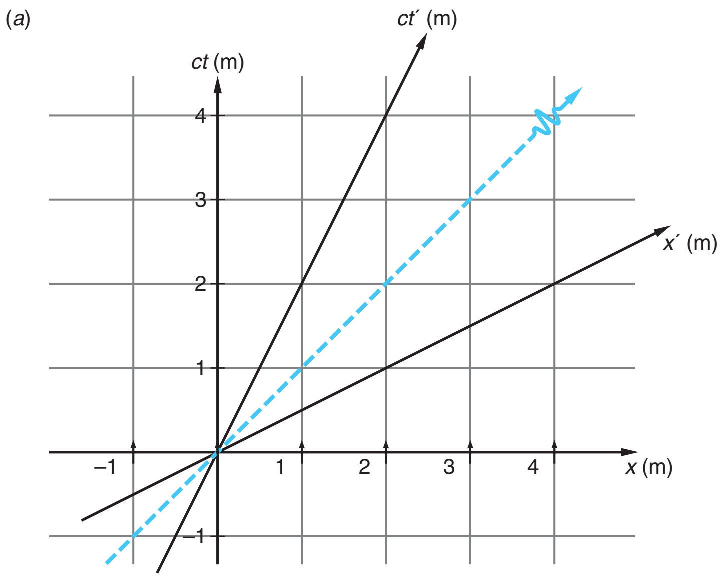
which says that the slope (in S) of the worldline of the point $x' = 0$, the ct' axis, is $1/\beta$. (See Figure 1-22*a*.)

In the same manner, the x' axis can be located using the fact that it is the locus of points for which $ct' = 0$. The Lorentz transformation once again provides the slope:

$$t' = \gamma\left(t - \frac{vx}{c^2}\right) = 0$$

$$\text{or} \qquad t = \frac{vx}{c^2} \qquad \text{and} \qquad ct = \frac{v}{c}x = \beta x$$

Thus, the slope of the x' axis as measured by an observer in S is β, as shown in Figure 1-22*a*. Don't be confused by the fact that the x axes don't look parallel anymore. They are still parallel in *space*, but this is a *spacetime* diagram. It shows motion in both



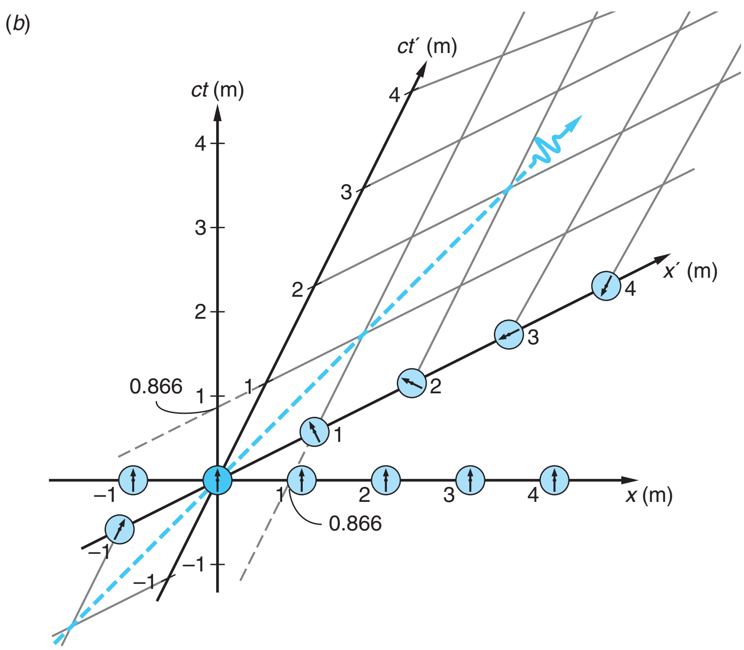


Figure 1-22 Spacetime diagram of S showing S' moving at speed $v = 0.5c$ in the $+x$ direction. The diagram is drawn with $t = t' = 0$ when the origins of S and S' coincided. The dashed line shows the worldline of a light flash that passed through the point $x = 0$ at $t = 0$ heading in the $+x$ direction. Its slope equals 1 in both S and S'. The ct' and x' axes of S' have slopes of $1/\beta = 2$ and $\beta = 0.5$, respectively. (*a*) Calibrating the axes of S' as described in the text allows the grid of coordinates to be drawn on S'. Interpretation is facilitated by remembering that (*b*) shows the system S' as it is observed in the spacetime diagram of S.

space and time. For example, the clock at $x' = 1$ m in Figure 1-22*b* passed the point $x = 0$ at about $ct = -1.5$ m as the x' axis of S' moved both upward and to the right in S. Remember, as time advances, the array of synchronized clocks and measuring rods that are the x axis also move upward, so that, for example, when $ct = 1$, the origin of S' ($x' = 0$, $ct' = 0$) has moved $vt = (v/c)ct = \beta ct$ to the right along the x axis.

Question

7. Explain how the spacetime diagram in Figure 1-22*b* would appear drawn by an observer in S'.

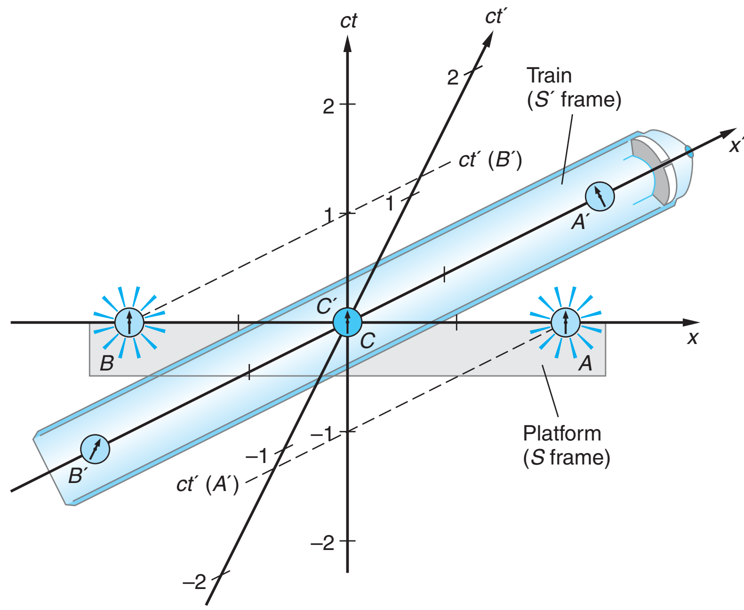


Figure 1-23 Spacetime equivalent of Figure 1-15, showing the spacetime diagram for the system S in which the platform is at rest. Measurements made by observers in S' are read from the primed axes.

EXAMPLE 1-6 Simultaneity in Spacetime Use the train-platform example of Figure 1-15 and a suitable spacetime diagram to show that events simultaneous in one frame are not simultaneous in a frame moving relative to the first. (This is the corollary to the relativity of simultaneity that we first demonstrated in the previous section using Figure 1-15.)

SOLUTION

Suppose a train is passing a station platform at speed v and an observer C at the mid-point of the platform, system S, announces that light flashes will be emitted at clocks A and B located at opposite ends of the platform at $t = 0$. Let the train, system S', be a rocket train with $v = 0.5c$. As in the earlier discussion, clocks at C and C' both read 0 as C' passes C. Figure 1-23 shows this situation. It is the spacetime equivalent of Figure 1-15.

Two events occur, the light flashes. The flashes are simultaneous in S since both occur at $ct = 0$. In S', however, the event at A occurred at $ct'(A')$ (see Figure 1-23), about 1.2 ct' units *before* $ct' = 0$, and the event at B occurred at $ct'(B')$, about 1.2 ct' units *after* $ct' = 0$. Thus, the flashes are not simultaneous in S' and A occurs before B, as we also saw in Figure 1-15.

EXPLORING
Calibrating the Spacetime Axes

By calibrating the coordinate axes of S' consistent with the Lorentz transformation, we will be able to read the coordinates of events and calculate space and time intervals between events as measured in both S and S' directly from the diagram, in addition to calculating them from Equations 1-18 and 1-19. The calibration of the S' axes is straightforward and is accomplished as follows. The locus of points, e.g., with $x' = 1$ m, is a line parallel to the ct' axis through the point $x' = 1$ m, $ct' = 0$, just as we saw earlier that the ct' axis was the locus of those points with $x' = 0$ through the point $x' = 0$, $ct' = 0$. Substituting these values into the Lorentz transformation for x', we see that the line through $x' = 1$ m intercepts the x axis, i.e., the line where $ct = 0$ at

$$\begin{aligned} x' &= \gamma(x - vt) = \gamma(x - \beta ct) \\ 1 &= \gamma x \qquad \text{or} \qquad x = 1/\gamma = \sqrt{1 - \beta^2} \end{aligned} \tag{1-24}$$

or, in general,

$$x = x'\sqrt{1 - \beta^2}$$

In Figure 1-22*b*, where $\beta = 0.5$, the line $x' = 1$ m intercepts the x axis at $x =$ 0.866 m. Similarly, if $x' = 2$ m, $x = 1.73$ m; if $x' = 3$ m, $x = 2.60$ m; and so on.

The ct' axis is calibrated in a precisely equivalent manner. The locus of points with $ct' = 1$ m is a line parallel to the x' axis through the point $ct' = 1$ m, $x' = 0$. Using the Lorentz transformation, the intercept of that line with the ct axis (where $x = 0$) is found as follows:

$$t' = \gamma(t - vx/c^2)$$

which can also be written as

$$ct' = \gamma(ct - \beta x) \tag{1-25}$$

or $ct' = \gamma ct$ for $x = 0$. Thus, for $ct' = 1$ m, we have $1 = \gamma ct$ or ct $= (1 - \beta^2)^{1/2}$ and, again in general, $ct = ct'(1 - \beta^2)^{1/2}$. The $x' \cdot ct'$ coordinate grid is shown in Figure 1-22*b*.

Notice in Figure 1-22*b* that the clocks located in S' are *not* found to be synchronized by observers in S, even though they are synchronized in S'. This is exactly the conclusion that we arrived at in the discussion of the lightning striking the train and platform. In addition, those with positive x' coordinates are behind the S' reference clock and those with negative x' coordinates are ahead, the differences being greatest for those clocks farthest away. This is a direct consequence of the Lorentz transformation of the time coordinate—i.e., when $ct = 0$ in Equation 1-25, $ct' = -\gamma\beta x$. Note, too, that the slope of the worldline of the light beam equals 1 in S' as well as in S, as required by the second postulate.

1-4 Time Dilation and Length Contraction

The results of correct measurements of the time and space intervals between events do not depend upon the kind of apparatus used for the measurements or on the events themselves. We are free therefore to choose any events and measuring apparatus that will help us understand the application of the Einstein postulates to the results of measurements. As you have already seen from previous examples, convenient events in relativity are those that produce light flashes. A convenient, simple such clock is a *light clock*, pictured schematically in Figure 1-24. A photocell detects the light pulse and sends a voltage pulse to an oscilloscope, which produces a vertical deflection of the oscilloscope's trace. The phosphorescent material on the face of the oscilloscope tube gives a persistent light that can be observed visually, photographed, or recorded electronically. The time between two light flashes is determined by measuring the

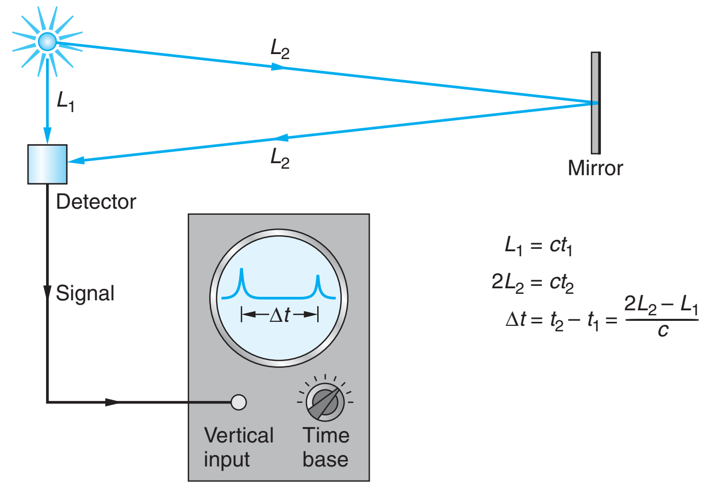


Figure 1-24 Light clock for measuring time intervals. The time is measured by reading the distance between pulses on the oscilloscope after calibrating the sweep speed.

distance between pulses on the scope and knowing the sweep speed. Such clocks can easily be calibrated and compared with other types of clocks. Although not drawn as in Figure 1-24, the clocks used in explanations in this section may be thought of as light clocks.

Time Dilation (or Time Stretching)

We first consider an observer A' at rest in frame S' a distance D from a mirror, also in S', as shown in Figure 1-25*a*. She triggers a flash gun and measures the time interval $\Delta t'$ between the original flash and the return flash from the mirror. Since light travels with speed c, this time is $\Delta t' = (2D)/c$.

Now consider these same two events, the original flash of light and the returning flash, as observed in reference frame S, with respect to which S' is moving to the right with speed v. The events happen at two different places, x_1 and x_2, in frame S because between the original flash and the return flash observer A' has moved a horizontal distance $v\Delta t$, where Δt is the time interval between the events measured in S.

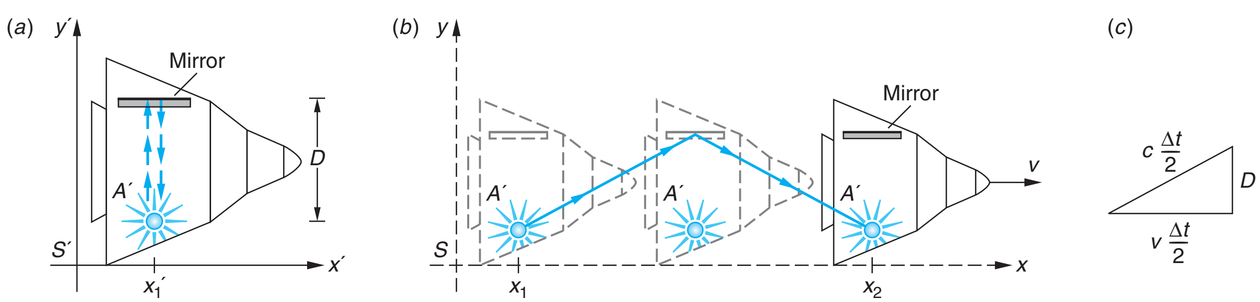


Figure 1-25 (*a*) Observer A' and the mirror are in a spaceship at rest in frame S'. The time it takes for the light pulse to reach the mirror and return is measured by A' to be $2D/c$. (*b*) In frame S, the spaceship is moving to the right with speed v. If the speed of light is the same in both frames, the time it takes for the light to reach the mirror and return is longer than $2D/c$ in S because the distance traveled is greater than $2D$. (*c*) A right triangle for computing the time Δt in frame S.

In Figure 1-25*b*, a space diagram, we see that the path traveled by the light is longer in S than in S'. However, by Einstein's postulates, light travels with the same speed c in frame S as it does in frame S'. Since it travels farther in S at the same speed, it takes longer in S to reach the mirror and return. The time interval between flashes in S is thus longer than it is in S'. We can easily calculate Δt in terms of $\Delta t'$. From the triangle in Figure 1-25*c*, we see that

$$\left(\frac{c\Delta t}{2}\right)^2 = D^2 + \left(\frac{v\Delta t}{2}\right)^2$$

or

$$\Delta t = \frac{2D}{\sqrt{c^2 - v^2}} = \frac{2D}{c}\frac{1}{\sqrt{1 - v^2/c^2}}$$

Using $\Delta t' = 2D/c$, we have

$$\Delta t = \frac{\Delta t'}{\sqrt{1 - v^2/c^2}} = \gamma\Delta t' = \gamma\tau \qquad \textbf{1-26}$$

where $\tau = \Delta t'$ is the *proper time interval* that we first encountered in Example 1-3. Equation 1-26 describes *time dilation*; i.e., it tells us that the observer in frame S always measures the time interval between two events to be longer (since $\gamma > 1$) than the corresponding interval measured on the clock located at both events in the frame where they occur at the same location. Thus, observers in S conclude that the clock at A' in S' runs slow since that clock measures a smaller time interval between the two events. Notice that the faster S' moves with respect to S, the larger is γ, and the slower the S' clocks will tick. It appears to the S observer that time is being stretched out in S′.

Be careful! The *same* clock must be located at each event for $\Delta t'$ to be the proper time interval τ. We can see why this is true by noting that Equation 1-26 can be obtained directly from the inverse Lorentz transformation for t. Referring again to Figure 1-25 and calling the emission of the flash event 1 and its return event 2, we have that

$$\Delta t = t_2 - t_1 = \gamma\left(t_2' + \frac{vx_2'}{c^2}\right) - \gamma\left(t_1' + \frac{vx_1'}{c^2}\right)$$

$$\Delta t = \gamma(t_2' - t_1') + \frac{\gamma v}{c^2}(x_2' - x_1')$$

or

$$\Delta t = \gamma\Delta t' + \frac{\gamma v}{c^2}\Delta x' \qquad \textbf{1-27}$$

If the clock that records t_2' and t_1' is located at the events, then $\Delta x' = 0$. If that is not the case, however, $\Delta x' \neq 0$ and $\Delta t'$, though certainly a valid measurement, is not a proper time interval. Only a clock located *at* an event *when* it occurs can record a proper time interval.

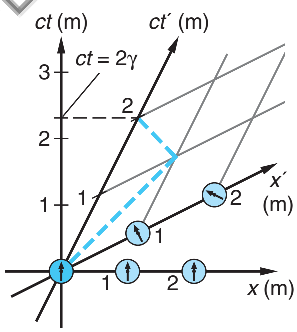


Figure 1-26 Spacetime diagram illustrating time dilation. The dashed line is the worldline of a light flash emitted at $x' = 0$ and reflected back to that point by a mirror at $x' = 1$ m. $\beta = 0.5$.

EXAMPLE 1-7 Spatial Separation of Events Two events occur at the same point x'_0 at times t'_1 and t'_2 in S', which moves with speed v relative to S. What is the spatial separation of these events measured in S?

SOLUTION

1. The location of the events in S is given by the Lorentz inverse transformation Equation 1-19:

$$x = \gamma(x' + vt')$$

2. The spatial separation of the two events $\Delta x = x_2 - x_1$ is then

$$\Delta x = \gamma(x'_0 + vt'_2) - \gamma(x'_0 + vt'_1)$$

3. The $\gamma x'_0$ terms cancel:

$$\Delta x = \gamma v(t'_2 - t'_1) = \gamma v \Delta t'$$

4. Since $\Delta t'$ is the proper time interval τ, Equation 1-26 yields

$$\Delta x = v\gamma\tau = v\Delta t$$

5. Using the situation in Figure 1-26 as a numerical example, where $\beta = 0.5$ and $\gamma = 1.15$, we have

$$\Delta x = \gamma \frac{v}{c}\Delta(ct') = (1.15)(0.5)(2)$$
$$= 1.15 \text{ m}$$

EXAMPLE 1-8 The Pregnant Elephant[14] Elephants have a gestation period of 21 months. Suppose that a freshly impregnated elephant is placed on a spaceship and sent toward a distant space jungle at $v = 0.75c$. If we monitor radio transmissions from the spaceship, how long after launch might we expect to hear the first squealing trumpet from the newborn calf?

SOLUTION

1. In S', the rest frame of the elephant, the time interval from launch to birth, is $\tau = 21$ months. In the Earth frame S the time interval is Δt_1, given by Equation 1-26:

$$\Delta t_1 = \gamma\tau = \frac{1}{\sqrt{1-\beta^2}}\tau$$
$$= \frac{1}{\sqrt{1-(0.75)^2}}(21 \text{ months})$$
$$= 31.7 \text{ months}$$

2. At that time the radio signal announcing the happy event starts toward Earth at speed c, but from where? Using the result of Example 1-7, since launch the spaceship has moved Δx in S, given by

$$\Delta x = \gamma vt = \gamma\beta ct$$
$$= (1.51)(0.75)(21\ c \cdot \text{months})$$
$$= 23.8\ c \cdot \text{months}$$

where $c \cdot$ month is the distance light travels in one month.

3. Notice that there is no need to convert Δx into meters since our interest is in how long it will take the radio signal to travel this distance in S. That time is Δt_2, given by

$$\Delta t_2 = \Delta x / c$$
$$= 23.8\ c \cdot \text{months}/c$$
$$= 23.8 \text{ months}$$

4. Thus, the good news will arrive at Earth at time Δt after launch where

$$\Delta t = \Delta t_1 + \Delta t_2$$
$$= 31.7 + 23.8$$
$$= 55.5 \text{ months}$$

Remarks: *This result, too, is readily obtained from a spacetime diagram. Figure 1-27 illustrates the general appearance of the spacetime diagram for this example, showing the elephant's worldline and the worldline of the radio signal.*

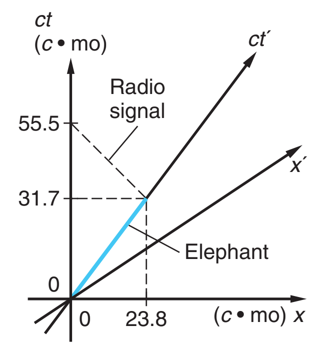


Figure 1-27 Sketch of the spacetime diagram for Example 1-8. $\beta = 0.75$. The colored line is the worldline of the pregnant elephant. The worldline of the radio signal is the dashed line at 45° toward the upper left.

Question

8. You are standing on a corner and a friend is driving past in an automobile. Both of you note the times when the car passes two different intersections and determine from your watch readings the time that elapses between the two events. Which of you has determined the proper time interval?

The time dilation of Equation 1-26 is easy to see in a spacetime diagram such as Figure 1-26, using the same round trip for a light pulse used above. Let the light flash leave $x' = 0$ at $ct' = 0$ when the S and S' origins coincided. The flash travels to $x' = 1$ m, reflects from a mirror located there, and returns to $x' = 0$. Let $\beta = 0.5$. The dotted line shows the worldline of the light beam, reflecting at ($x' = 1$, $ct' = 1$) and returning to $x' = 0$ at $ct' = 2$ m. Note that the S observer records the latter event at $ct > 2$ m; i.e., the observer in S sees the S' clock running slow.

Experimental tests of the time dilation prediction have been performed using macroscopic clocks, in particular, accurate atomic clocks. In 1975, C. O. Alley conducted a test of both general and special relativity in which a set of atomic clocks were carried by a U.S. Navy antisubmarine patrol aircraft while it flew back and forth over the same path for 15 hours at altitudes between 8000 m and 10,000 m over Chesapeake Bay. The clocks in the plane were compared by laser pulses with an identical group of clocks on the ground. (See Figure 1-13 for one way such a comparison might be done.) Since the experiment was primarily intended to test the gravitational effect on clocks predicted by general relativity (see Section 2-5), the aircraft was deliberately flown at the rather sedate average speed of 270 knots (140 m/s) = $4.7 \times 10^{-7}c$ to minimize the time dilation due to the relative speeds of the clocks. Even so, after Alley deducted the effect of gravitation as predicted by general relativity, the airborne clocks lost an average of 5.6×10^{-9} s due to the relative speed during the 15-hour flight. This result agrees with the prediction of special relativity, 5.7×10^{-9} s, to within 2 percent, even at this low relative speed. The experimental results leave little basis for further debate as to whether traveling clocks of all kinds lose time on a round trip. They do.

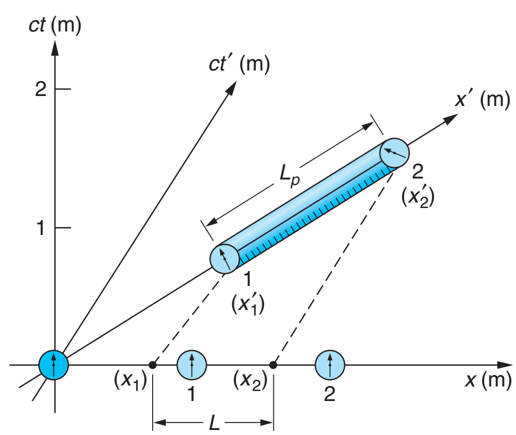


Figure 1-28 A measuring rod, a meter stick in this case, lies at rest in S' between $x_2' = 2$ m and $x_1' = 1$ m. System S' moves with $\beta = 0.79$ relative to S. Since the rod is in motion, S must measure the locations of the ends of the rod x_2 and x_1 simultaneously in order to have made a valid length measurement. L is obviously shorter than L_p. By direct measurement from the diagram (use a millimeter scale) $L/L_p = 0.61 = 1/\gamma$.

Length Contraction

A phenomenon closely related to time dilation is *length contraction*. The length of an object measured in the reference frame in which the object is at rest is called its *proper length* L_p. In a reference frame in which the object is moving, the measured length parallel to the direction of motion is shorter than its proper length. Consider a rod at rest in the frame S' with one end at x_2' and the other end at x_1', as illustrated in Figure 1-28. The length of the rod in this frame is its proper length $L_p = x_2' - x_1'$. Some care must be taken to find the length of the rod in frame S. In this frame, the rod is moving to the right with speed v, the speed of frame S'. The length of the rod in frame S is *defined* as $L = x_2 - x_1$, where x_2 is the position of one end at some time t_2 and x_1 is the position of the other end *at the same time* $t_1 = t_2$ as measured in frame S. Since the rod is at rest in S', t'_2 need not equal t'_1. Equation 1-18 is convenient to use to calculate $x_2 - x_1$ at some time t because it relates x, x', and t, whereas Equation 1-19 is not convenient because it relates x, x', and t':

$$x'_2 = \gamma(x_2 - vt_2) \qquad \text{and} \qquad x'_1 = \gamma(x_1 - vt_1)$$

Since $t_2 = t_1$, we obtain

$$x'_2 - x'_1 = \gamma(x_2 - x_1)$$

$$x_2 - x_1 = \frac{1}{\gamma}(x'_2 - x'_1) = \sqrt{1 - \frac{v^2}{c^2}}(x'_2 - x'_1)$$

or

$$L = \frac{1}{\gamma}L_p = \sqrt{1 - \frac{v^2}{c^2}}L_p \qquad \textbf{1-28}$$

Thus, the length of a rod is smaller when it is measured in a frame with respect to which it is moving. Before Einstein's paper was published, Lorentz and G. FitzGerald had independently shown that the null result of the Michelson-Morley experiment could be explained by assuming that the lengths in the direction of the interferometer's motion contracted by the amount given in Equation 1-28. For that reason, the length contraction is often called the *Lorentz-FitzGerald contraction*.

EXAMPLE 1-9 Speed of S' A stick that has a proper length of 1 m moves in a direction parallel to its length with speed v relative to you. The length of the stick as measured by you is 0.914 m. What is the speed v?

SOLUTION

1. The length of the stick measured in a frame relative to which it is moving with speed v is related to its proper length by Equation 1-28:

$$L = \frac{L_p}{\gamma}$$

2. Rearranging to solve for γ:

$$\gamma = \frac{L_p}{L}$$

3. Substituting the values of L_p and L:

$$\gamma = \frac{1\text{ m}}{0.914\text{ m}} = \frac{1}{\sqrt{1 - v^2/c^2}}$$

4. Solving for v:

$$\sqrt{1 - v^2/c^2} = 0.914$$
$$1 - v^2/c^2 = (0.914)^2 = 0.835$$
$$v^2/c^2 = 1 - 0.835 = 0.165$$
$$v^2 = 0.165c^2$$
$$v = 0.406c$$

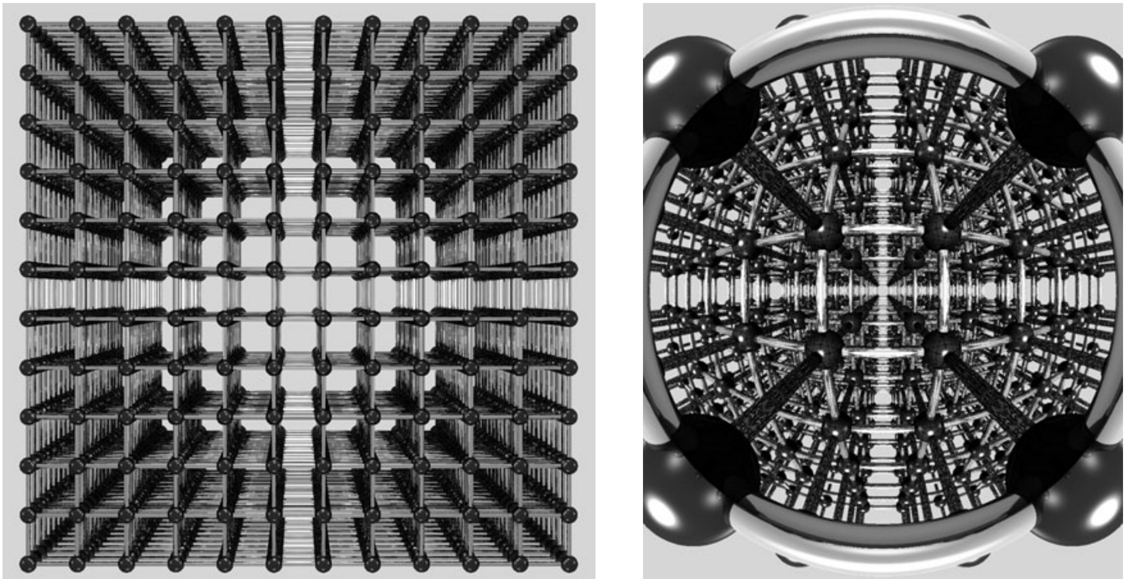


Figure 1-29 The appearance of rapidly moving objects depends on both length contraction in the direction of motion and the time when the observed light left the object. (*a*) The array of clocks and measuring rods that represents S' as viewed by an observer in S with $\beta = 0$. (*b*) When S' approaches the S observer with $\beta = 0.9$, the distortion of the lattice becomes apparent. This is what an observer on a cosmic-ray proton might see as it passes into the lattice of a face-centered-cubic crystal such as NaCl. [*P.-K. Hsiung, R. Dunn, and C. Cox. Courtesy of C. Cox, Adobe Systems, Inc., San Jose, CA.*]

It is important to remember that the relativistic contraction of moving lengths occurs only parallel to the relative motion of the reference frames. In particular, observers in relatively moving systems measure the same values for lengths in the y and y' and in the z and z' directions perpendicular to their relative motion. The result is that observers measure different shapes and angles for two- and three-dimensional objects. (See Example 1-10 and Figures 1-29 and 1-30.)

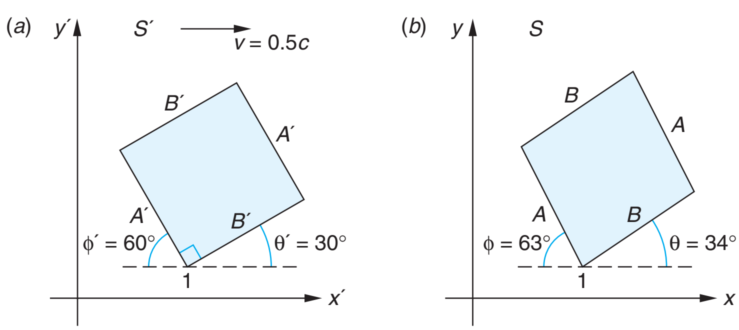


Figure 1-30 Length contraction distorts the shape and orientation of two- and three-dimensional objects. The observer in S measures the square shown in S' as a rotated parallelogram.

EXAMPLE 1-10 The Shape of a Moving Square Consider the square in the $x'y'$ plane of S' with one side making a 30° angle with the x' axis, as in Figure 1-30a. If S' moves with $\beta = 0.5$ relative to S, what is the shape and orientation of the figure in S?

SOLUTION

The S observer measures the x components of each side to be shorter by a factor $1/\gamma$ than those measured in S'. Thus, S measures

$$A = [\cos^2 30 + \sin^2 30/\gamma^2]^{1/2} A' = 0.968 A'$$

$$B = [\sin^2 30 + \cos^2 30/\gamma^2]^{1/2} B' = 0.901 B'$$

Since the figure is a square in S', $A' = B'$. In addition, the angles between B and the x axis and between A and the x axis are given by, respectively,

$$\theta = \tan^{-1}\left[\frac{B' \sin 30}{B' \cos 30/\gamma}\right] = \tan^{-1}\left[\gamma \frac{\sin 30}{\cos 30}\right] = 33.7°$$

$$\phi = \tan^{-1}\left[\frac{A' \cos 30}{A' \sin 30/\gamma}\right] = \tan^{-1}\left[\gamma \frac{\cos 30}{\sin 30}\right] = 63.4°$$

Thus, S concludes from geometry that the interior angle at vertex 1 is not 90°, but 180° − (63.4° + 33.7°) = 82.9°—i.e., the figure is not a square, but a parallelogram whose shorter sides make 33.7° angles with the x axis! Its shape and orientation in S are shown in Figure 1-30b.

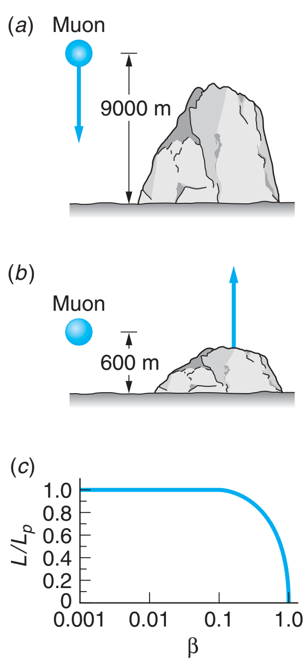


Figure 1-31 Although muons are created high above Earth and their mean lifetime is only about 2 μs when at rest, many appear at Earth's surface. (a) In Earth's reference frame, a typical muon moving at 0.998c has a mean lifetime of 30 μs and travels 9000 m in this time. (b) In the reference frame of the muon, the distance traveled by Earth is only 600 m in the muon's lifetime of 2 μs. (c) L varies only slightly from L_p until v is of the order of 0.1c. $L \to 0$ as $v \to c$.

Muon Decay

An interesting example of both time dilation and length contraction is afforded by the appearance of muons as secondary radiation from cosmic rays. Muons decay according to the statistical law of radioactivity:

$$N(t) = N_0 e^{(-t/\tau)} \qquad \textbf{1-29}$$

where N_0 is the original number of muons at time $t = 0$, $N(t)$ is the number remaining at time t, and τ is the mean lifetime (a proper time interval), which is about 2 μs for muons. Since muons are created (from the decay of pions) high in the atmosphere, usually several thousand meters above sea level, few muons should reach sea level. A typical muon moving with speed 0.998c would travel only about 600 m in 2 μs. However, the lifetime of the muon measured in Earth's reference frame is increased according to time dilation (Equation 1-26) by the factor $1/(1 - v^2/c^2)^{1/2}$, which is 15 for this particular speed. The mean lifetime measured in Earth's reference frame is therefore 30 μs, and a muon with speed 0.998c travels about 9000 m in this time. From the muon's point of view, it lives only 2 μs, but the atmosphere is rushing past it with a speed of 0.998c. The distance of 9000 m in Earth's frame is thus contracted to only 600 m in the muon's frame, as indicated in Figure 1-31.

It is easy to distinguish experimentally between the classical and relativistic predictions of the observations of muons at sea level. Suppose that we observe 10^8 muons at an altitude of 9000 m in some time interval with a muon detector. How many would we expect to observe at sea level in the same time interval? According to the nonrelativistic prediction, the time it takes for these muons to travel 9000 m is (9000 m)/$0.998c \approx 30\ \mu s$, which is 15 lifetimes. Substituting $N_0 = 10^8$ and $t = 15\tau$ into Equation 1-29, we obtain

$$N = 10^8 e^{-15} = 30.6$$

We would thus expect all but about 31 of the original 100 million muons to decay before reaching sea level.

According to the relativistic prediction, Earth must travel only the contracted distance of 600 m in the rest frame of the muon. This takes only $2\ \mu s = 1\tau$. Therefore, the number of muons expected at sea level is

$$N = 10^8 e^{-1} = 3.68 \times 10^7$$

Thus relativity predicts that we would observe 36.8 million muons in the same time interval. Experiments of this type have confirmed the relativistic predictions.

The Spacetime Interval

We have seen earlier in this section that time intervals and lengths (= space intervals), quantities that were absolutes, or invariants, for relatively moving observers using the classical Galilean coordinate transformation, are not invariants in special relativity. The Lorentz transformation and the relativity of simultaneity lead observers in inertial frames to conclude that lengths moving relative to them are contracted and time intervals are stretched, both by the factor γ. The question naturally arises: Is there *any* quantity involving the space and time coordinates that is invariant under a Lorentz transformation? The answer to that question is yes, and as it happens, we have already dealt with a special case of that invariant quantity when we first obtained the correct form of the Lorentz transformation. It is called the *spacetime interval*, or usually just the *interval*, Δs, and is given by

Experiments with muons moving near the speed of light are performed at many accelerator laboratories throughout the world despite their short mean life. Time dilation results in much longer mean lives relative to the laboratory, providing plenty of time to do experiments.

$$(\Delta s)^2 = (c\Delta t)^2 - [\Delta x^2 + \Delta y^2 + \Delta z^2] \qquad \textbf{1-30}$$

or, specializing it to the one-space-dimensional systems that we have been discussing,

$$(\Delta s)^2 = (c\Delta t)^2 - (\Delta x)^2 \qquad \textbf{1-31}$$

It may help to think of Equations 1-30 and 1-31 like this:

$$[\text{interval}]^2 = [\text{separation in time}]^2 - [\text{separation in space}]^2$$

The interval Δs is the only measurable quantity describing pairs of events in spacetime for which observers in all inertial frames will obtain the same numerical value. The negative sign in Equations 1-30 and 1-31 implies that $(\Delta s)^2$ may be positive, negative, or zero depending on the relative sizes of the time and space separations. With the sign of $(\Delta s)^2$, nature is telling us about the causal relation between the two events. Notice that whichever of the three possibilities characterizes a pair for one observer, it does so for all observers since Δs is invariant. The interval is called *timelike* if the time separation is the larger and *spacelike* if the space separation predominates. If the two terms are equal, so that $\Delta s = 0$, then it is called *lightlike*.

Timelike Interval Consider a material particle[15] or object, e.g., the elephant in Figure 1-27, that moves relative to S. Since no material particle has ever been measured traveling faster than light, particles always travel less than 1 m of distance in 1 m of light travel time. We saw that to be the case in Example 1-8, where the time interval between launch and birth of the baby was 31.7 months on the S clock, during which time the elephant had moved a distance of $23.8c \cdot$ months. Equation 1-31 then yields $(c\Delta t)^2 - (\Delta x)^2 = (31.7c)^2 - (23.8c)^2 = (21.0c)^2 = (\Delta s)^2$, and the interval in S is $\Delta s = 21.0\ c \cdot$ months. The time interval term being the larger, Δs is a timelike interval and we say that material particles have *timelike worldlines*. Such worldlines lie within the shaded area of the spacetime diagram in Figure 1-21. Note that in the elephant's frame S' the separation in space between the launch and birth is zero and Δt is 21.0 months. Thus $\Delta s = 21.0\ c \cdot$ months in S', too. That is what we mean by the interval being invariant: observers in both S and S' measure the same number for the separation of the two events in spacetime.

The proper time interval τ between two events can be determined from Equations 1-31 using space and time measurements made in *any* inertial frame since we can write that equation as

$$\frac{\Delta s}{c} = \sqrt{(\Delta t)^2 - (\Delta x/c)^2}$$

Since $\Delta t = \tau$ when $\Delta x = 0$—i.e., for the time interval recorded on a clock in a system moving such that the clock is located at each event as it occurs—in that case

$$\sqrt{(\Delta t)^2 - (\Delta x/c)^2} = \sqrt{\tau^2 - 0} = \tau = \frac{\Delta s}{c} \qquad \textbf{1-32}$$

Notice that this yields the correct proper time $\tau = 21.0$ months in the elephant example.

Spacelike Interval When two events are separated in space by an interval whose square is greater than the value of $(c\Delta t)^2$, then Δs is called *spacelike*. In that case it is convenient for us to write Equation 1-31 in the form

$$(\Delta s)^2 = (\Delta x)^2 - (c\Delta t)^2 \qquad \textbf{1-33}$$

so that, as with timelike intervals, $(\Delta s)^2$ is not negative.[16] Events that are spacelike occur sufficiently far apart in space and close together in time that no inertial frame could move fast enough to carry a clock from one event to the other. For example, suppose two observers in Earth frame S, one in San Francisco and one in London, agree to each generate a light flash at the same instant, so that $c\Delta t = 0$ m in S and $\Delta x = 1.08 \times 10^7$ m. For *any* other inertial frame $(c\Delta t)^2 > 0$, and we see from Equation 1-33 that $(\Delta x)^2$ must be greater than $(1.08 \times 10^7)^2$ in order that Δs be invariant. In other words, 1.08×10^7 m is as close in space as the two events can be in any system; consequently, it will not be possible to find a system moving fast enough to move a clock from one event to the other. A speed greater than c, in this case infinitely greater, would be needed. Notice that the value of $\Delta s = L_p$, the proper length. Just as with the proper time interval τ, measurements of space and time intervals in any inertial system can be used to determine L_p.

Lightlike (or Null) Interval The relation between two events is *lightlike* if Δs in Equation 1-31 equals zero. In that case

$$c\Delta t = \Delta x \tag{1-34}$$

and a light pulse that leaves the first event as it occurs will just reach the second as it occurs.

The existence of the lightlike interval in relativity has no counterpart in the world of our everyday experience, where the geometry of space is Euclidean. In order for the distance between two points in space to be zero, the separation of the points in each of the three space dimensions must be zero. However, in spacetime the interval between two events may be zero, even though the intervals in space and time may individually be quite large. Notice, too, that pairs of events separated by lightlike intervals have both the proper time interval and proper length equal to zero since $\Delta s = 0$.

Things that move at the speed of light[17] have lightlike worldlines. As we saw earlier (see Figure 1-22), the worldline of light bisects the angles between the ct and x axes in a spacetime diagram. Timelike intervals lie in the shaded areas of Figure 1-32 and share the common characteristic that their relative order in time is the same for observers in all inertial systems. Events A and B in Figure 1-32 are such a pair. Observers in both S and S' agree that A occurs *before* B, although they of course measure different values for the space and time separations. Causal events, i.e., events that depend upon or affect one another in some fashion, such as your birth and that of your mother, have timelike intervals. On the other hand, the temporal order of events with spacelike intervals, such as A and C in Figure 1-32, depends upon the relative motion of the systems. As you can see in the diagram, A occurs before C in S, but C occurs first in S'. Thus, the relative order of pairs of events is absolute in the shaded areas but elsewhere may be in either order.

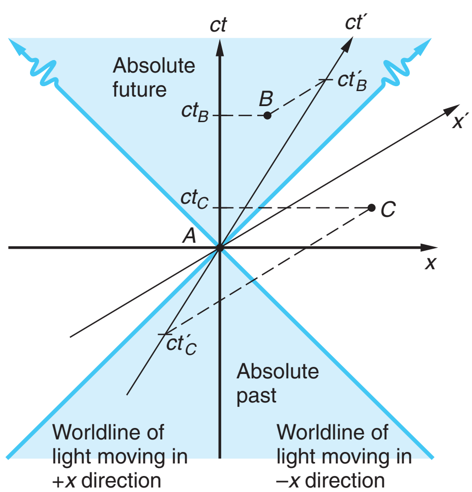


Figure 1-32 The relative temporal order of events for pairs characterized by timelike intervals, such as A and B, is the same for all inertial observers. Events in the upper shaded area will all occur in the future of A; those in the lower shaded area occurred in A's past. Events whose intervals are spacelike, such as A and C, can be measured as occurring in either order, depending on the relative motion of the frames. Thus, C occurs after A in S but before A in S'.

Question

9. In 1987 light arrived at Earth from the explosion of a star (a supernova) in the Large Magellanic Cloud, a small companion galaxy to the Milky Way, located about 170,000 $c \cdot$ y away. Describe events that together with the explosion of the star would be separated from it by (a) a spacelike interval, (b) a lightlike interval, and (c) a timelike interval.

EXAMPLE 1-11 Characterizing Spacetime Intervals Figure 1-33 is the spacetime diagram of a laboratory showing three events, the emission of light from an atom in each of three samples.

1. Determine whether the interval between each of the three possible pairs of events is timelike, spacelike, or lightlike.
2. Would it have been possible in any of the pairs for one of the events to have been caused by the other? If so, which?

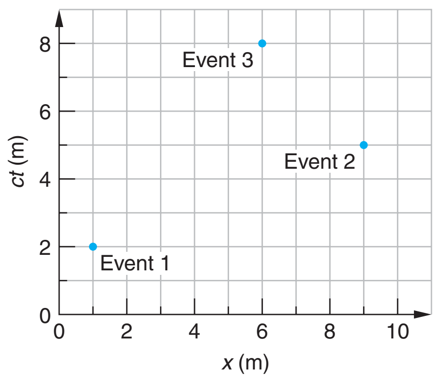


Figure 1-33 A spacetime diagram of three events whose intervals Δs are found in Example 1-11.

SOLUTION

1. The spacetime coordinates of the events are

event	ct	x
1	2	1
2	5	9
3	8	6

and for the three possible pairs 1 and 2, 2 and 3, and 1 and 3 we have

pair	$c\Delta t$	Δx	$(c\Delta t)^2$	$(\Delta x)^2$	
1 & 2	5–2	9–1	9	64	spacelike
2 & 3	8–5	6–9	9	9	lightlike
1 & 3	8–2	6–1	36	25	timelike

2. Yes, event 3 may possibly have been caused by either event 1 since 3 is in the absolute future of 1, or event 2, since 2 and 3 can just be connected by a flash of light.

1-5 The Doppler Effect

In the Doppler effect for sound the change in frequency for a given velocity v depends on whether it is the source or receiver that is moving with that speed. Such a distinction is possible for sound because there is a medium (the air) relative to which the motion takes place, and so it is not surprising that the motion of the source or the receiver relative to the still air can be distinguished. Such a distinction between motion of the source or receiver cannot be made for light or other electromagnetic waves in a vacuum as a consequence of Einstein's second postulate; therefore, the classical expressions for the Doppler effect cannot be correct for light. We will now derive the relativistic Doppler effect equations that are correct for light.

Consider a light source moving toward an observer or receiver at A in Figure 1-34a at velocity v. The source is emitting a train of light waves toward receivers A and B while approaching A and receding from B. Figure 1-34b shows the spacetime

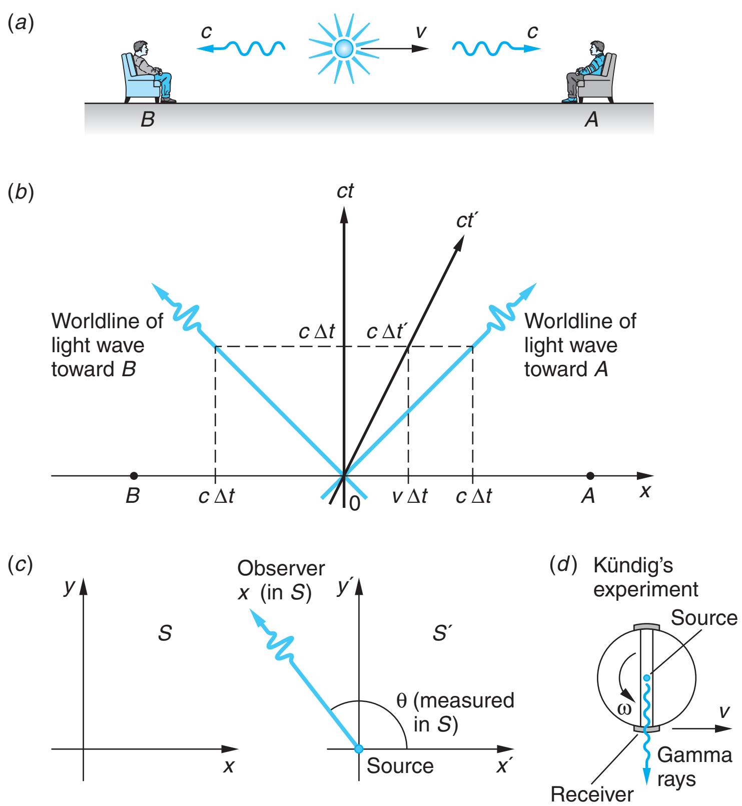


Figure 1-34 The Doppler effect in light, as in sound, arises from the relative motion of the source and receiver; however, the independence of the speed of light on that motion leads to different expressions for the frequency shift. (a) A source approaches observer A and recedes from observer B. The spacetime diagram of the system S in which A and B are at rest and the source moves at velocity $\mathbf{v}$ illustrates the two situations. (b) The source located at $x' = 0$ (the x' axis is omitted) moves along its worldline, the ct' axis. The N waves emitted toward A in time Δt occupy space $\Delta x = c\Delta t - v\Delta t$, whereas those headed for B occupy $\Delta x = c\Delta t + v\Delta t$. In three dimensions the observer in S may see light emitted at some angle θ with respect to the x axis as in (c). In that case a transverse Doppler effect occurs. (d) Kündig's apparatus for measuring the transverse Doppler effect.

diagram of S, the system in which A and B are at rest. The source is located at $x' = 0$ (x' axis is not shown), and, of course, its worldline is the ct' axis. Let the source emit a train of N electromagnetic waves in each direction beginning when the S and S' origins were coincident. First, let's consider the train of waves headed toward A. During the time Δt over which the source emits the N waves, the first wave emitted will have traveled a distance $c\Delta t$ and the source itself a distance $v\Delta t$ in S. Thus, the N waves are seen by the observer at A to occupy a distance $c\Delta t - v\Delta t$ and, correspondingly, their wavelength λ is given by

$$\lambda = \frac{c\Delta t - v\Delta t}{N}$$

and the frequency $f = c/\lambda$ is

$$f = \frac{c}{\lambda} = \frac{cN}{(c - v)\Delta t} = \frac{1}{1 - \beta}\frac{N}{\Delta t}$$

The frequency of the source in S', called the *proper frequency*, is given by $f_0 = c/\lambda' = N/\Delta t'$, where $\Delta t'$ is measured in S', the rest system of the source. The time interval $\Delta t' = \tau$ is the proper time interval since the light waves, in particular the first and the Nth, are all emitted at $x' = 0$; hence $\Delta x' = 0$ between the first and the Nth in S'. Thus, Δt and $\Delta t'$ are related by Equation 1-26 for time dilation, so $\Delta t = \gamma\Delta t'$, and when the source and receiver are moving toward each other, the observer A in S measures the frequency

$$f = \frac{1}{1 - \beta}\frac{f_0 \Delta t'}{\Delta t} = \frac{f_0}{1 - \beta}\frac{1}{\gamma} \qquad \textbf{1-35}$$

or

$$f = \frac{\sqrt{1 - \beta^2}}{1 - \beta} f_0 = \sqrt{\frac{1 + \beta}{1 - \beta}} f_0 \qquad \text{(approaching)} \qquad \textbf{1-36}$$

This differs from the classical equation only in the addition of the time dilation factor. Note that $f > f_0$ for the source and observer approaching each other. Since for visible light this corresponds to a shift toward the blue part of the spectrum, it is called a *blueshift*.

Suppose the source and receiver are moving away from each other, as for observer B in Figure 1-34*b*. Observer B, in S, sees the N waves occupying a distance $c\Delta t + v\Delta t$, and the same analysis shows that observer B in S measures the frequency

$$f = \frac{\sqrt{1 - \beta^2}}{1 + \beta} f_0 = \sqrt{\frac{1 - \beta}{1 + \beta}} f_0 \qquad \text{(receding)} \qquad \textbf{1-37}$$

The use of Doppler radar to track weather systems is a direct application of special relativity.

Notice that $f < f_0$ for the observer and source receding from each other. Since for visible light this corresponds to a shift toward the red part of the spectrum, it is called a *redshift*. It is left as a problem for you to show that the same results are obtained when the analysis is done in the frame in which the source is at rest.

In the event that $v \ll c$ (i.e., $\beta \ll 1$), as is often the case for light sources moving on Earth, useful (and easily remembered) approximations of Equations 1-36 and 1-37 can be obtained. Using Equation 1-36 as an example and rewriting it in the form

$$f = f_0(1 + \beta)^{1/2}(1 - \beta)^{-1/2}$$

the two quantities in parentheses can be expanded by the binomial theorem to yield

$$f = f_0\left(1 + \frac{1}{2}\beta - \frac{1}{8}\beta^2 + \ldots\right)\left(1 + \frac{1}{2}\beta + \frac{3}{8}\beta^2 + \ldots\right)$$

Multiplying out and discarding terms of higher order than β yields

$$f/f_0 \approx 1 + \beta \qquad \text{(approaching)}$$

and, similarly,

$$f/f_0 \approx 1 - \beta \qquad \text{(receding)}$$

and $|\Delta f/f_0| \approx \beta$ in both situations, where $\Delta f = f_0 - f$.

EXAMPLE 1-12 Rotation of the Sun The Sun rotates at the equator once in about 25.4 days. The Sun's radius is 7.0×10^8 m. Compute the Doppler effect that you would expect to observe at the left and right limbs (edges) of the Sun near the equator for light of wavelength $\lambda = 550$ nm $= 550 \times 10^{-9}$ m (yellow light). Is this a redshift or a blueshift?

SOLUTION

The speed of limbs v = (circumference)/(time for one revolution) or

$$v = \frac{2\pi R}{T} = \frac{2\pi(7.0 \times 10^8)\text{ m}}{25.4\text{ d} \cdot 3600\text{ s/h} \cdot 24\text{ h/d}} = 2000\text{ m/s}$$

$v \ll c$, so we may use the approximation equations. Using $\Delta f/f_0 \approx \beta$, we have that $\Delta f \approx \beta f_o = \beta c/\lambda_o$, or $\Delta f \approx 2000/550 \times 10^{-9} = 3.64 \times 10^9$ Hz. Since $f_o = c/\lambda_o = (3 \times 10^8\text{ m/s})/(550 \times 10^{-9}) = 5.45 \times 10^{14}$ Hz, Δf represents a fractional change in frequency of β, or about one part in 10^5. It is a redshift for the receding limb, a blueshift for the approaching one.

Doppler Effect of Starlight

In 1929 E. P. Hubble became the first astronomer to suggest that the universe is expanding.[18] He made that suggestion and offered a simple equation to describe the expansion on the basis of measurements of the Doppler shift of the frequencies of light emitted toward us by distant galaxies. Light from distant galaxies is always shifted toward frequencies lower than those emitted by similar sources nearby. Since the general expression connecting the frequency f and wavelength λ of light is $c = f\lambda$, the shift corresponds to longer wavelengths. As noted above, the color red is on the longer-wavelength side of the visible spectrum (see Chapter 4), so the *redshift* is used to describe the Doppler effect for a receding source. Similarly, *blueshift* describes light emitted by stars, typically stars in our galaxy, that are approaching us.

Astronomers define the redshift of light from astronomical sources by the expression $z = (f_o - f)/f$, where f_o = frequency measured in the frame of the star or galaxy and f = frequency measured at the receiver on Earth. This allows us to write $\beta = v/c$ in terms of z as

$$\beta = \frac{(z + 1)^2 - 1}{(z + 1)^2 + 1} \qquad \textbf{1-38}$$

Equation 1-37 is the appropriate one to use for such calculations, rather than the approximations, since galactic recession velocities can be quite large. For example, the quasar 2000-330 has a measured $z = 3.78$, which implies from Equation 1-38 that it is receding from Earth due to the expansion of space at $0.91c$. (See Chapter 13.)

EXAMPLE 1-13 Redshift of Starlight The longest wavelength of light emitted by hydrogen in the Balmer series (see Chapter 4) has a wavelength of $\lambda_o = 656$ nm. In light from a distant galaxy, this wavelength is measured as $\lambda = 1458$ nm. Find the speed at which the galaxy is receding from Earth.

SOLUTION

1. The recession speed is the v in $\beta = v/c$. Since $\lambda > \lambda_o$, this is a redshift and Equation 1-37 applies:

$$f = \sqrt{\frac{1-\beta}{1+\beta}}f_0$$

2. Rewriting Equation 1-37 in terms of the wavelengths:

$$f = \sqrt{\frac{1-\beta}{1+\beta}} = \frac{f}{f_0} = \frac{\lambda_o}{\lambda}$$

3. Squaring both sides and substituting values for λ_o and λ:

$$\frac{1-\beta}{1+\beta} = \left(\frac{\lambda_0}{\lambda}\right)^2$$

$$= \left(\frac{656 \text{ nm}}{1458 \text{ nm}}\right)^2 = 0.202$$

4. Solving for β:

$$1 - \beta = (0.202)(1 + \beta)$$

$$1.202\beta = 1 - 0.202 = 0.798$$

$$\beta = \frac{0.798}{1.202} = 0.664$$

5. The galaxy is thus receding at speed v, where

$$v = c\beta = 0.664c$$

EXPLORING
Transverse Doppler Effect

Our discussion of the Doppler effect in Section 1-5 involved only one space dimension, wherein the source, observer, and the direction of the relative motion all lie on the x axis. In three space dimensions, where they may not be colinear, a more complete analysis, though beyond the scope of our discussion, makes only a small change in Equation 1-35. If the source moves along the positive x axis but the observer views the light emitted at some angle θ with the x axis, as shown in Figure 1-34c, Equation 1-35 becomes

$$f = \frac{f_0}{\gamma}\frac{1}{1 - \beta\cos\theta} \qquad \textbf{1-35}\boldsymbol{a}$$

When $\theta = 0$, this becomes the equation for the source and receiver approaching, and when $\theta = \pi$, the equation becomes that for the source and receiver receding. Equation 1-35a also makes the quite surprising prediction that even when viewed perpendicular to the direction of motion, where $\theta = \pi/2$, the observer will still see a frequency shift, the so-called *transverse Doppler effect*, $f = f_0/\gamma$. Note that $f < f_0$ since $\gamma > 1$. It is sometimes referred to as the second-order Doppler effect and is the result of time dilation of the moving source. [The general derivation of Equation 1-35*a* can be found in the French (1968), Resnick (1992), and Ohanian (2001) references at the end of the chapter.]

Following a suggestion first made by Einstein in 1907, Kündig in 1962 made an excellent quantitative verification of the transverse Doppler effect.[19] He used 14.4-keV gamma rays emitted by a particular isotope of Fe as the light source (see Chapter 11). The source was at rest in the laboratory, on the axis of an ultracentrifuge, and the receiver (an Fe absorber foil) was mounted on the ultracentrifuge rim, as shown in Figure 1-34*d*. Using the extremely sensitive frequency measuring technique called the Mössbauer effect (see Chapter 11), Kündig found a transverse Doppler effect in agreement with the relativistic prediction within ± 1 percent over a range of relative speeds up to about 400 m/s.

1-6 The Twin Paradox and Other Surprises

The consequences of Einstein's postulates—the Lorentz transformation, relativistic velocity addition, time dilation, length contraction, and the relativity of simultaneity—lead to a large number of predictions that are unexpected and even startling when compared with our experiences in a macroscopic world where $\beta \approx 0$ and geometry obeys the Euclidean rules. Still other predictions seem downright paradoxical, with relatively moving observers obtaining equally valid but apparently totally inconsistent results. This chapter concludes with the discussion of a few such examples that will help you hone your understanding of special relativity.

Twin Paradox

Perhaps the most famous of the paradoxes in special relativity is that of the twins or, as it is sometimes called, the clock paradox. It arises out of time dilation (Equation 1-26) and goes like this. Homer and Ulysses are identical twins. Ulysses travels at a constant high speed to a star beyond our solar system and returns to Earth while his twin, Homer, remains at home. When the traveler Ulysses returns home, he finds his twin brother much aged compared to himself—in agreement, we will see, with the prediction of relativity. The paradox arises out of the contention that the motion is relative and either twin could regard the other as the traveler, in which case each twin should find the other to be younger than he and we have a logical contradiction—a paradox. Let's illustrate the paradox with a specific example. Let Earth and the destination star be in the same inertial frame S. Two other frames S' and S'' move relative to S at $v = +0.8c$ and $v = -0.8c$, respectively. Thus, $\gamma = 5/3$ in both cases. The spaceship carrying Ulysses accelerates quickly from S to S', then coasts with S' to the star, again accelerates quickly from S' to S'', coasts with S'' back to Earth, and brakes to a stop alongside Homer.

It is easy to analyze the problem from Homer's point of view on Earth. Suppose, according to Homer's clock, Ulysses coasts in S' for a time interval $\Delta t = 5$ y and in S'' for an equal time. Thus, Homer is 10 y older when Ulysses returns. The time interval in S' between the events of Ulysses' leaving Earth and arriving at the star is shorter because it is a proper time interval. The time it takes to reach the star by Ulysses' clock is

$$\Delta t' = \frac{\Delta t}{\gamma} = \frac{5 \text{ y}}{5/3} = 3 \text{ y}$$

Since the same time is required for the return trip, Ulysses will have recorded 6 y for the round trip and will be 4 y younger than Homer upon his return.

The difficulty in this situation seems to be for Ulysses to understand why his twin aged 10 y during his absence. If we consider Ulysses as being at rest and Homer as moving away, Homer's clock should run slow and measure only $3/\gamma = 1.8$ y, and it appears that Ulysses should expect Homer to have aged only 3.6 years during the round trip. This is, of course, the paradox. Both predictions can't be right. However, this approach makes the incorrect assumption that the twins' situations are symmetrical and interchangeable. They are not. Homer remains in a single inertial frame, whereas Ulysses *changes* inertial frames, as illustrated in Figure 1-35*a*, the spacetime diagram for Ulysses' trip. While the turnaround may take only a minute fraction of the total time, it is absolutely essential if the twins clocks are to come together again so that we can compare their ages (readings).

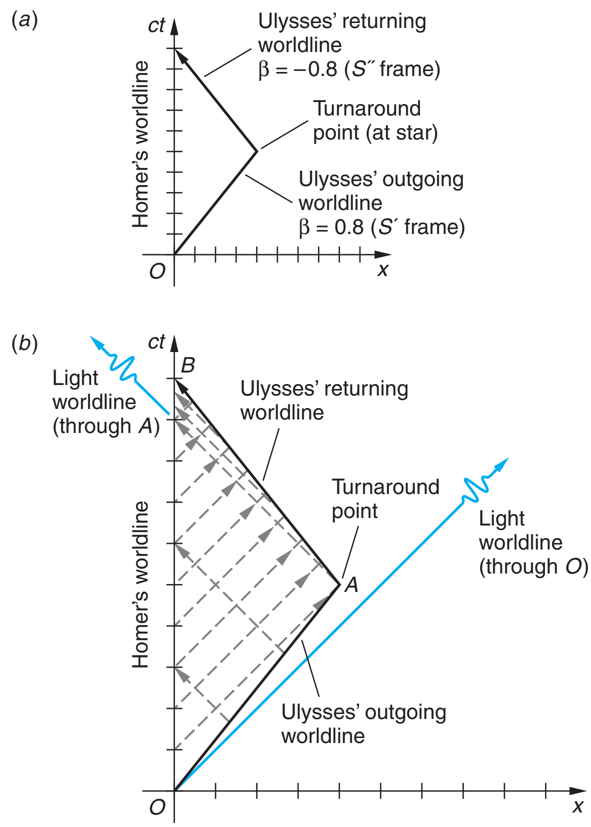


Figure 1-35 (*a*) The spacetime diagram of Ulysses' journey to a distant star in the inertial frame in which Homer and the star are at rest. (*b*) Divisions on the *ct* axis correspond to years on Homer's clock. The broken lines show the paths (worldlines) of light flashes transmitted by each twin with a frequency of 1/year on his clock. Note the markedly different frequencies at the receivers.

A correct analysis can be made using the invariant interval Δs from Equation 1-31 rewritten as

$$\left(\frac{\Delta s}{c}\right)^2 = (\Delta t)^2 - \left(\frac{\Delta x}{c}\right)^2$$

where the left side is constant and equal to $(\tau)^2$, the proper time interval squared, and the right side refers to measurements made in any inertial frame. Thus, Ulysses along each of his worldlines in Figure 1-35*a* has $\Delta x = 0$ and, of course, measures $\Delta t = \tau = 3$ y, or 6 y for the round trip. Homer, on the other hand, measures

$$(\Delta t)^2 = (\tau)^2 + \left(\frac{\Delta x}{c}\right)^2$$

and since $(\Delta x/c)^2$ is always positive, he always measures $\Delta t > \tau$. In this situation $\Delta x = 0.8c\Delta t$, so

$$(\Delta t)^2 = (3\text{ y})^2 + (0.8c\Delta t/c)^2$$

or

$$(\Delta t)^2(0.36) = (3)^2$$

$$\Delta t = \frac{3}{0.6} = 5\text{ y}$$

or 10 y for the round trip, as we saw earlier. The reason that the twins' situations cannot be treated symmetrically is because the special theory of relativity can predict the behavior of accelerated systems, such as Ulysses at the turnaround, provided that in the formulation of the physical laws we take the view of an inertial, i.e., unaccelerated, observer such as Homer. That's what we have done. Thus, we cannot do the same analysis in the rest frame of Ulysses' spaceship because it does not remain in an inertial frame during the round trip; hence, it falls outside of the special theory, and no paradox arises. The laws of physics can be reformulated so as to be invariant for accelerated observers, which is the role of general relativity (see Chapter 2), but the result is the same: Ulysses returns younger than Homer by just the amount calculated above.

EXAMPLE 1-14 Twin Paradox and the Doppler Effect This example, first suggested by C. G. Darwin,[20] may help you understand what each twin sees during Ulysses' journey. Homer and Ulysses agree that once each year, on the anniversary of the launch date of Ulysses' spaceship (when their clocks were together), each twin will send a light signal to the other. Figure 1-35*b* shows the light signals each sends. Based on our discussion above, Homer sends 10 light flashes (the *ct* axis, Homer's worldline, is divided into 10 equal interval corresponding to the 10 years of the journey on Homer's clock) and Ulysses sends 6 light flashes (each of Ulysses' worldlines is divided into 3 equal intervals corresponding to 3 years on Ulysses' clock). Note that each transmits his final light flash as they are reunited at *B*. Although each transmits light signals with a frequency of 1 per year, they obviously do not receive them at that frequency. For example, Ulysses sees no signals from Homer during the first three years! How can we explain the observed frequencies?

SOLUTION

The Doppler effect provides the explanation. As the twins (and clocks) recede from each other the frequency of their signals is reduced from the proper frequency f_0 according to Equation 1-37 and we have

$$\frac{f}{f_0} = \sqrt{\frac{1-\beta}{1+\beta}} = \sqrt{\frac{1-0.8}{1+0.8}} = \frac{1}{3}$$

which is exactly what both twins see (refer to Figure 1-35*b*): Homer receives 3 flashes in the first 9 years and Ulysses 1 flash in his first 3 years; i.e., $f = (1/3)f_0$ for both.

After the turnaround they are approaching each other and Equation 1-38 yields

$$\frac{f}{f_0} = \sqrt{\frac{1+\beta}{1-\beta}} = \sqrt{\frac{1+0.8}{1-0.8}} = 3$$

and again this agrees with what the twins see: Homer receives 3 flashes during the final (10th) year and Ulysses receives 9 flashes during his final 3 years; i.e., $f = 3f_o$ for both.

Question

10. The different ages of the twins upon being reunited are an example of the relativity of simultaneity that was discussed earlier. Explain how that accounts for the fact that their biological clocks are no longer synchronized.

More

It is the relativity of simultaneity, not their different accelerations, that is responsible for the age difference between the twins. This is readily illustrated in *The Case of the Identically Accelerated Twins*, which can be found on the home page: www.whfreeman.com/tiplermodern-physics5e. See also Figure 1-36 here.

The Pole and Barn Paradox

An interesting problem involving length contraction, reported initially by W. Rindler, involves putting a long pole into a short barn. One version, from E. F. Taylor and J. A. Wheeler,[22] goes as follows. A runner carries a pole 10 m long toward the open front door of a small barn 5 m long. A farmer stands near the barn so that he can see both the front and the back doors of the barn, the latter being a closed swinging door, as shown in Figure 1-37*a*. The runner carrying the pole at speed v enters the barn, and at some instant the farmer sees the pole completely contained in the barn and closes the

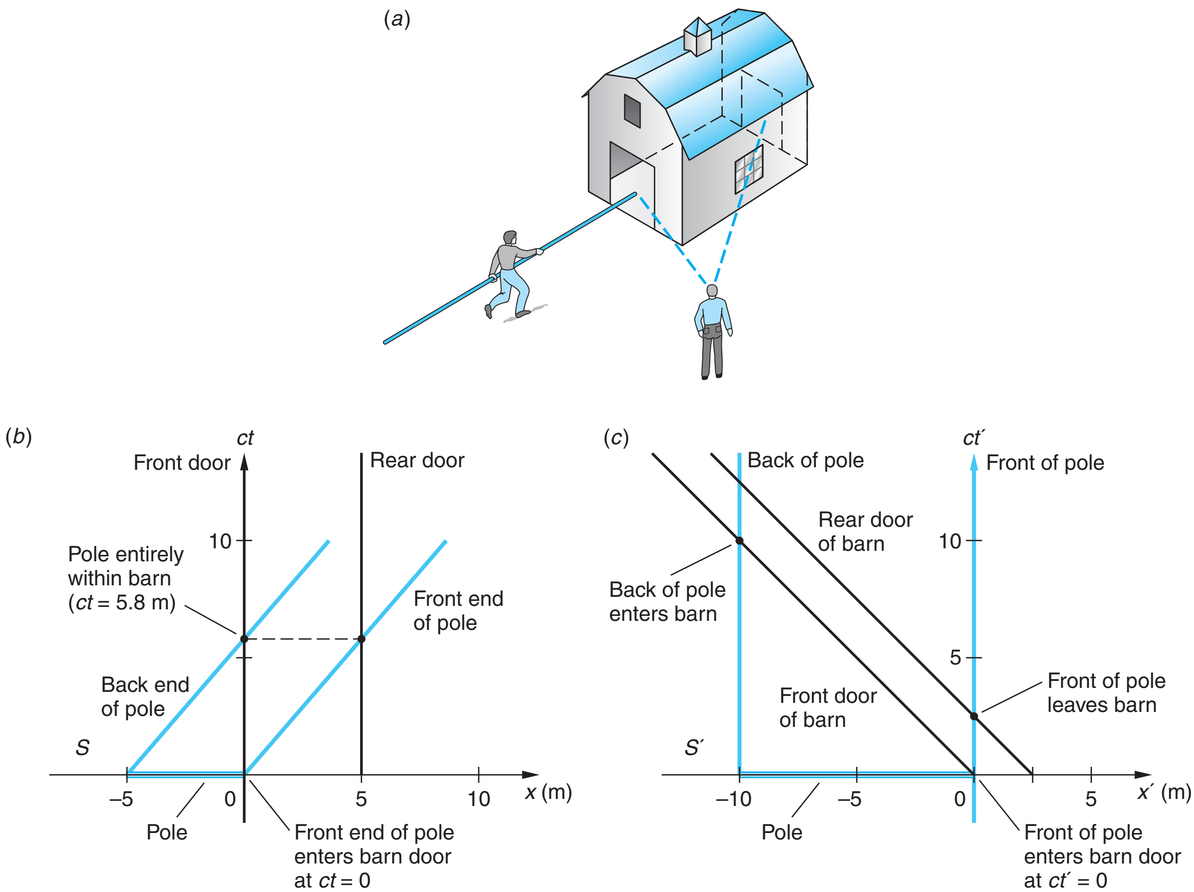


Figure 1-37 (*a*) A runner carrying a 10-m pole moves quickly enough so that the farmer will see the pole entirely contained in the barn. The spacetime diagrams from the point of view of the farmer's inertial frame (*b*) and that of the runner (*c*). The resolution of the paradox is in the fact that the events of interest, shown by the large dots in each diagram, are simultaneous in S but not in S'.

front door, thus putting a 10-m pole into a 5-m barn. The minimum speed of the runner v that is necessary for the farmer to accomplish this feat can be computed from Equation 1-28, giving the relativistic length contraction $L = L_p/\gamma$, where L_p = proper length of the pole (10 m) and L = length of the pole measured by the farmer, to be equal to the length of the barn (5 m). Therefore, we have

$$\gamma = \frac{1}{\sqrt{1 - v^2/c_2}} = \frac{L_p}{L} = \frac{10}{5}$$

$$1 - v^2/c^2 = (5/10)^2$$

$$v^2/c^2 = 1 - (5/10)^2 = 0.75$$

$$v = 0.866c \quad \text{or} \quad \beta = 0.866$$

A paradox seems to arise when this situation is viewed in the rest system of the runner. For him the pole, being at rest in the same inertial system, has its proper length of 10 m. However, the runner measures the length of the barn to be

$$L = L_p/\gamma = 5\sqrt{1 - \beta^2}$$
$$L = 2.5 \text{ m}$$

How can he possibly fit the 10-m pole into the length-contracted 2.5-m barn? The answer is that he can't, and the paradox vanishes, but how can that be? To understand the answer, we need to examine two events—the coincidences of both the front and back ends of the pole, respectively, with the rear and front doors of the barn—in the inertial frame of the farmer and in that of the runner.

These are illustrated by the spacetime diagrams of the inertial frame S of the farmer and barn (Figure 1-37*b*) and that of the runner S' (Figure 1-37*c*). Both diagrams are drawn with the front end of the pole coinciding with the front door of the barn at the instant the clocks are started. In Figure 1-37*b* the worldlines of the barn doors are, of course, vertical, while those of the two ends of the pole make an angle $\theta = \tan^{-1}(1/\beta) = 49.1°$ with the x axis. Note that in S the front of the pole reaches the rear door of the barn at $ct = 5 \text{ m}/0.866 = 5.8 \text{ m}$ *simultaneous* with the arrival of the back end of the pole at the front door; i.e., at that instant in S the pole is entirely contained in the barn.

In the runner's rest system S' it is the worldlines of the ends of the pole that are vertical, while those of the front and rear doors of the barn make angles of 49.1° with the $-x'$ axis (since the barn moves in the $-x'$ direction at v). Now we see that the rear door passes the front of the pole at $ct' = 2.5 \text{ m}/0.866 = 2.9 \text{ m}$, but the front door of the barn doesn't reach the rear of the pole until $ct' = 10 \text{ m}/0.866 = 11.5 \text{ m}$. Thus, the first of those two events occurs *before* the second, and the runner never sees the pole entirely contained in the barn. Once again, the relativity of simultaneity is the key—events simultaneous in one inertial frame are not necessarily simultaneous when viewed from another inertial frame.

Now let's consider a different version of this paradox, the one initially due to W. Rindler. Suppose the barn's back wall were made of thick, armor-plate steel and had no door. What do the farmer and the runner see then? Once again, in the farmer's (and the barn's) rest frame, the instant the front of the pole reaches the armor plate, the farmer shuts the door and the 10-m pole is instantaneously contained in the 5-m barn. However, in the *next* instant (assuming that the pole doesn't break) it must either bend (i.e., rotate in spacetime) or break through the armor plate. Since this is relativity, the runner must come to the same conclusion in his rest frame as the 2.5-m barn races toward him at $\beta = 0.866$. But now when the armor plate back wall contacts the front of the pole, the barn continues to move at $\beta = 0.866$, taking the front of the pole with it and leaving at that instant 7.5 m of the pole still outside the barn. Yet like the farmer, the runner must also see the 10-m pole entirely contained within the 2.5-m barn. How can that be? Like this: the instant the tip of the pole hits the steel plate, that information (an elastic shock wave) begins to propagate down the pole. Even if the wave were to propagate at the speed of light c, it will take $10 \text{ m}/3.0 \times 10^8 \text{ m/s} = 3.33 \times 10^{-8} \text{ s}$ to reach the back of the pole. In the meantime, the barn door must move only 7.5 m to reach the back of the pole and does so in only $7.5 \text{ m}/(0.866 \times 3.0 \times 10^8 \text{ m/s}) = 2.89 \times 10^{-8} \text{ s}$. Thus, the runner, in agreement with the farmer, sees the 10-m pole entirely contained within the 2.5-m barn—at least briefly!

Question

11. In the discussion where the barn's back wall was made from armor-plate steel and had no door, do the farmer and the runner both see the pole entirely contained in the barn, no matter what their relative speed is? Explain.

Headlight Effect

We have made frequent use of Einstein's second postulate asserting that the speed of light is independent of the source motion for all inertial observers; however, the same is not true for the *direction* of light. Consider a light source in S' that emits light uniformly in all directions. A beam of that light emitted at an angle θ' with respect to the x' axis is shown in Figure 1-38*a*. During a time $\Delta t'$ the x' displacement of the beam is $\Delta x'$, and these are related to θ' by

$$\frac{\Delta x'}{c\Delta t'} = \frac{\Delta x'}{\Delta(ct')} = \cos\theta' \qquad \textbf{1-39}$$

The direction of the beam relative to the x axis in S is similarly given by

$$\frac{\Delta x}{\Delta(ct)} = \cos\theta \qquad \textbf{1-40}$$

Applying the inverse Lorentz transformation to Equation 1-40 yields

$$\cos\theta = \frac{\Delta x}{c\Delta t} = \frac{\gamma(\Delta x' + v\Delta t')}{c\gamma(\Delta t' + v\Delta x'/c^2)}$$

Dividing the numerator and denominator by $\Delta t'$ and then by c, we obtain

$$\cos\theta = \frac{(\Delta x'/\Delta t' + v)}{c\left(1 + \frac{v}{c^2}\Delta x'/\Delta t'\right)} = \frac{\Delta x'/\Delta(ct') + v/c}{1 + \frac{v}{c}\cdot\frac{\Delta x'}{\Delta(ct')}}$$

and substituting from Equation 1-39 yields

$$\cos\theta = \frac{\cos\theta' + \beta}{1 + \beta\cos\theta'} \qquad \textbf{1-41}$$

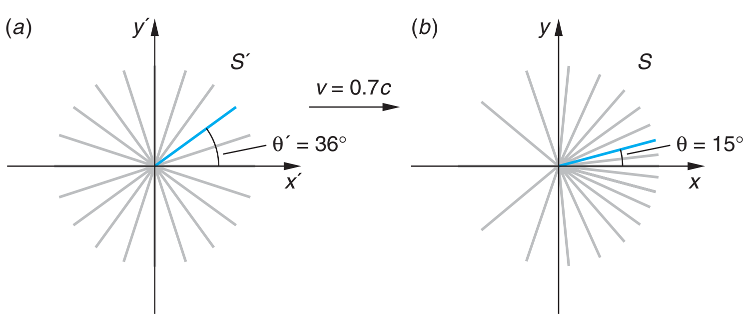


Figure 1-38 (*a*) The source at rest in S' moves with $\beta = 0.7$ with respect to S. (*b*) Light emitted uniformly in S' appears to S concentrated into a cone in the forward direction. Rays shown in (*a*) are 18° apart. Rays shown in (*b*) make angles calculated from Equation 1-41. The two colored rays shown are corresponding ones.

Considering the half of the light emitted by the source in S' into the forward hemisphere, i.e., rays with θ' between $\pm\pi/2$, note that Equation 1-41 restricts the angles θ measured in S for those rays (50 percent of all the light) to lie between $\theta = \pm\cos^{-1}\beta$. For example, for $\beta = 0.5$, the observer in S would see half of the total light emitted by the source in S' to lie between $\theta = \pm 60°$, i.e., in a cone of half angle 60° whose axis is along the direction of the velocity of the source. For values of β near unity θ is very small, e.g., $\beta = .99$ yields $\theta = 8.1°$. This means that the observer in S sees half of all the light emitted by the source to be concentrated into a forward cone with that half angle. (See Figure 1-38*b*.) Note, too, that the remaining 50 percent of the emitted light is distributed throughout the remaining 344° of the two-dimensional diagram.[23] As a result of the headlight effect, light from a directly approaching source appears far more intense than that from the same source at rest. For the same reason, light from a directly receding source will appear much dimmer than that from the same source at rest. This result has substantial applications in experimental particle physics and astrophysics.

In determining the brightness of stars and galaxies, a critical parameter in understanding them, astronomers must correct for the headlight effect, particularly at high velocities relative to Earth.

Question

12. Notice from Equation 1-41 that some light emitted by the moving source into the rear hemisphere is seen by the observer in S as having been emitted into the forward hemisphere. Explain how that can be, using physical arguments.

EXPLORING Superluminal Speeds

We conclude this chapter with a few comments about things that move faster than light. The Lorentz transformations (Equations 1-18 and 1-19) have no meaning in the event that the relative speeds of two inertial frames exceed the speed of light. This is generally taken to be a prohibition on the moving of mass, energy, and information faster than c. However, it is possible for certain processes to proceed at speeds greater than c and for the speeds of moving objects to appear to be greater than c without contradicting relativity theory. A common example of the first of these is the motion of the point where the blades of a giant pair of scissors intersect as the scissors are quickly closed, sometimes called the scissors paradox. Figure 1-39 shows the situation. A long straight rod (one blade) makes an angle θ with the x axis (the second blade) and moves in the $-y$ direction at constant speed $v_y = \Delta y/\Delta t$. During time Δt, the intersection of the blades, point P, moves to the right a distance Δx. Note from the figure that

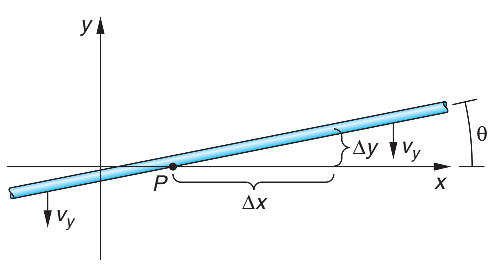


Figure 1-39 As the long straight rod moves vertically downward, the intersection of the "blades," point P, moves toward the right at speed $v_p = \Delta x/\Delta t$. In terms of v_y and θ, $v_p = v_y/\tan\theta$.

$\Delta y/\Delta x = \tan\theta$. The speed with which P moves to the right is

$$v_p = \Delta x/\Delta t = \frac{\Delta x}{\Delta y/v_y} = \frac{v_y \Delta x}{\Delta x \tan\theta} \qquad \textbf{1-42}$$

or

$$v_p = \frac{v_y}{\tan\theta}$$

Since $\tan\theta \to 0$ as $\theta \to 0$, it will always be possible to find a value of θ close enough to zero so that $v_p > c$ for any (nonzero) value of v_y. As real scissors are closed, the angle gets progressively smaller, so in principle all that one needs for $v_p > c$ are long blades so that $\theta \to 0$.

Question

13. Use a diagram like Figure 1-32 to explain why the motion of point P cannot be used to convey information to observers along the blades.

The point P in the scissors paradox is, of course, a geometrical point, not a material object, so it is perhaps not surprising that it could appear to move at speeds greater than c. As an example of an object with mass appearing to do so, consider a tiny meteorite moving through space directly toward you at high speed v. As it enters Earth's atmosphere, about 9 km above the surface, frictional heating causes it to glow and the first light from the glow starts toward your eye. After some time Δt the frictional heating has evaporated all of the meteorite's matter, the glow is extinguished, and its final light starts toward your eye, as illustrated in Figure 1-40. During the time between the first and the final glow, the meteorite traveled a distance $v\Delta t$. During that same time interval light from the first glow has traveled toward your eye a distance $c\Delta t$. Thus, the space interval between the first and final glows is given by

$$\Delta y = c\Delta t - v\Delta t = \Delta t(c - v)$$

and the visual time interval at your eye Δt_{eye} between the arrival of the first and final light is

$$\Delta t_{\text{eye}} = \Delta y/c = \frac{\Delta t(c - v)}{c} = \Delta t(1 - \beta)$$

and, finally, the apparent visual speed v_a that you record is

$$v_a = \frac{v\Delta t}{\Delta t_{\text{eye}}} = \frac{v\Delta t}{\Delta t(1 - \beta)} = \frac{\beta c}{1 - \beta} \qquad \textbf{1-43}$$

Clearly, $\beta = 0.5$ yields $v_a = c$ and any larger β yields $v_a > c$. For example, a meteorite approaching you at $v = 0.8c$ is perceived to be moving at $v_a = 4c$. Certain galactic

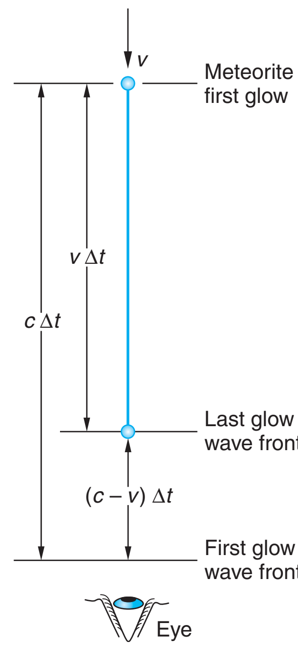


Figure 1-40 A meteorite moves directly toward the observer's eye at speed v. The spatial distance between the wave fronts is $(c - v)\Delta t$ as they move at c, so the time interval between their arrival at the observer is not Δt, but Δt_{eye}, which is $(c - v)\Delta t/c = (1 - \beta)\Delta t$, and the apparent speed of approach is $v_a = v\Delta t/\Delta t_{\text{eye}} = \beta c/(1 - \beta)$.

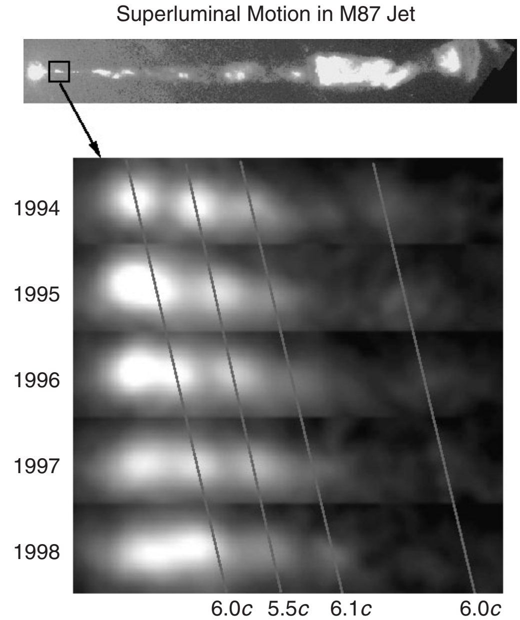


Figure 1-41 Superluminal motion has been detected in a number of cosmic objects. This sequence of images taken by the Hubble Space Telescope shows apparent motion at six times the speed of light in galaxy M87. The jet streaming from the galaxy's nucleus (the bright round region at the far left in the bar image at the top) is about 5000 $c \cdot$y long. The boxed region is enlarged. The slanting lines track the moving features and indicate the apparent speeds in each region. [*John Biretta, Space Telescope Science Institute.*]

structures may also be observed to move at superluminal speeds, as the sequence of images of the jet from galaxy M87 in Figure 1-41 illustrates.

As a final example of things that move faster than c, it has been proposed that particles with mass might exist whose speeds would always be faster than light speed. One basis for this suggestion is an appealing symmetry: ordinary particles always have $v < c$, and photons and other massless particles have $v = c$, so the existence of particles with $v > c$ would give a sort of satisfying completeness to the classification of particles. Called *tachyons*, their existence would present relativity with serious but not necessarily insurmountable problems of infinite creation energies and causality paradoxes, e.g., alteration of history. (See the next example.) No compelling theoretical arguments preclude their existence and eventual discovery; however, to date, all experimental searches for tachyons[24] have failed to detect them, and the limits set by those experiments indicate that it is highly unlikely they exist.

EXAMPLE 1-15 Tachyons and Reversing History Use tachyons and an appropriate spacetime diagram to show how the existence of such particles might be used to change history and, hence, alter the future, leading to a paradox.

SOLUTION

In a spacetime diagram of the laboratory frame S the worldline of a particle with $v > c$ created at the origin traveling in the $+x$ direction makes an angle less than 45° with the x axis; i.e., it is below the light worldline, as shown in Figure 1-42. After some time the tachyon reaches a tachyon detector mounted on a spaceship moving rapidly away at $v < c$ in the $+x$ direction. The spaceship frame S' is shown in the figure at P. The detector immediately creates a new tachyon, sending it off in the $-x'$ direction and, of course, into the future of S', i.e., with $ct' > 0$. The second tachyon returns to the laboratory at $x = 0$ but at a time ct *before* the first tachyon was emitted, having traveled into the past of S to point M, where $ct < 0$. Having sent an object into our own past, we would then have the ability to alter events that occur after M and produce causal contradictions. For example, the laboratory tachyon detector could be coupled to equipment that created the first tachyon via a computer programmed to

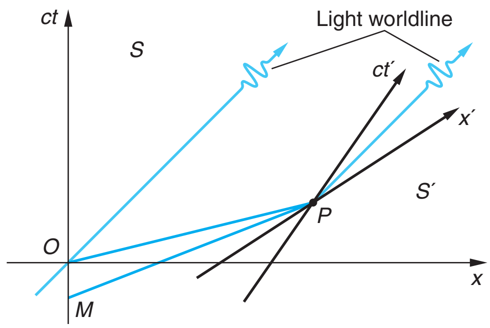


Figure 1-42 A tachyon emitted at O in S, the laboratory frame, catches up with a spaceship moving at high speed at P. Its detection triggers the emission of a second tachyon at P back toward the laboratory at $x = 0$. The second tachyon arrives at the laboratory at $ct < 0$, i.e., before the emission of the first tachyon.

cancel emission of the first tachyon if the second tachyon is detected. (Shades of *The Terminator*!) It is logical that contradictions such as this, together with the experimental results referred to above, lead to the conclusion that faster-than-light particles do not exist.

As mentioned above, one attraction (or specter) associated with objects moving faster than light is the prospect of altering history via time travel. We close this chapter on relativity by illustrating one such paradox in Figure 1-43.

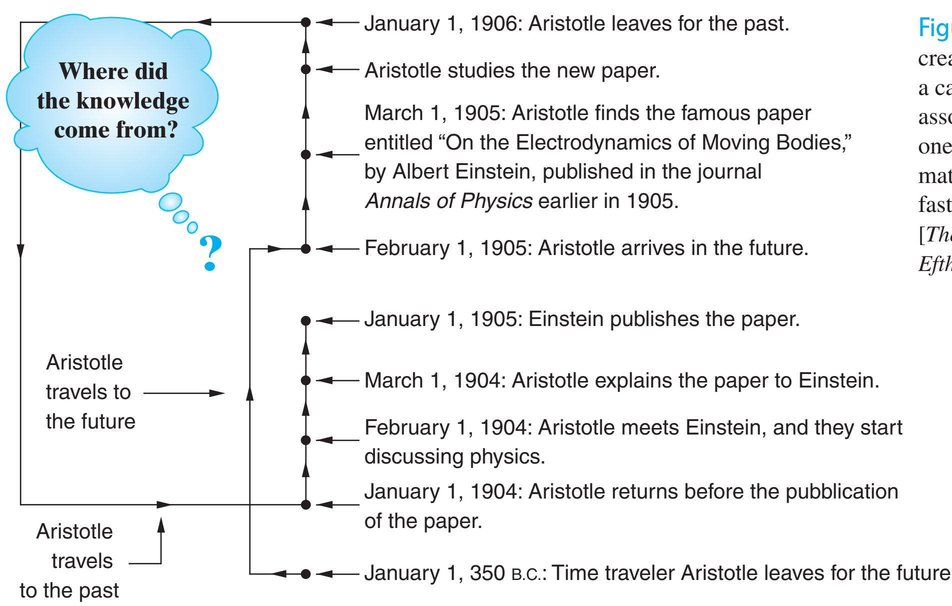


Figure 1-43 The knowledge creation paradox illustrates a causality problem associated with time travel, one possible consequence of material objects moving faster than light speed. [*The authors thank Costas Efthimiou for this example.*]

Summary

TOPIC	RELEVANT EQUATIONS AND REMARKS	
1. Classical relativity		
Galilean transformation	$x' = x - vt \qquad y' = y \qquad z' = z \qquad t' = t$	**1-2**
Newtonian relativity	Newton’s laws are invariant in all systems connected by a Galilean transformation.	

TOPIC	RELEVANT EQUATIONS AND REMARKS	
2. Einstein's postulates	The laws of physics are the same in all inertial reference frames. The speed of light is c, independent of the motion of the source.	
3. Relativity of simultaneity	Events simultaneous in one reference frame are not in general simultaneous in any other inertial frame.	
4. Lorentz transformation	$x' = \gamma(x - vt) y' = y\ z' = z$ $t' = \gamma(t - vx/c^2)$ with $\gamma = (1 - v^2/c^2)^{-1/2}$	**1-18**
5. Time dilation	Proper time is the time interval τ between two events that occur at the same space point. If that interval is $\Delta t' = \tau$, then the time interval in S is $\Delta t = \gamma \Delta t' = \gamma\tau$ where $\gamma = (1 - v^2/c^2)^{-1/2}$	**1-26**
6. Length contraction	The proper length of a rod is the length L_p measured in the rest system of the rod. In S, moving at speed v with respect to the rod, the length measured is $L = L_p/\gamma$	**1-28**
7. Spacetime interval	All observers in inertial frames measure the same interval Δs between pairs of events in spacetime, where $(\Delta s)^2 = (c\Delta t)^2 - (\Delta x)^2$	**1-31**
8. Doppler effect		
Source/ observer approaching	$f = \sqrt{\dfrac{1+\beta}{1-\beta}}\, f_0$	
Source/ observer receding	$f = \sqrt{\dfrac{1-\beta}{1+\beta}}\, f_0$	

General References

The following general references are written at a level appropriate for readers of this book.

Bohm, D., *The Special Theory of Relativity*, W. A. Benjamin, New York, 1965.

French, A. P., *Special Relativity*, Norton, 1968. Includes an excellent discussion of the historical basis of relativity.

Gamow, G., *Mr. Tompkins in Paperback*, Cambridge University Press, Cambridge, 1965. Contains the delightful Mr. Tompkins stories. In one of these Mr. Tompkins visits a dream world where the speed of light is only about 10 mi/h and relativistic effects are quite noticeable.

Lorentz, H. A., A. Einstein, H. Minkowski, and W. Weyl, *The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity* (trans. W. Perrett and J. B. Jeffery), Dover, New York, 1923. A delightful little book containing Einstein's original paper ["On the Electrodynamics of Moving Bodies," *Annalen der Physik*, **17** (1905)] and several other original papers on special relativity.

Ohanian, H. C., *Special Relativity: A Modern Introduction*, Physics Curriculum & Instruction, 2001.

Pais, A., *Subtle Is the Lord . . .* , Oxford University Press, Oxford, 1982.

Resnick, R., and D. Halliday, *Basic Concepts in Relativity and Early Quantum Theory,* 2d ed., Macmillan, 1992.

Rindler, W., *Essential Relativity*, Van Nostrand Reinhold, New York, 1969.

Taylor, E. F., and J. A. Wheeler, *Spacetime Physics,* 2d ed., W. H. Freeman and Co. 1992. This is a good book with many examples, problems, and diagrams.

Notes

1. Polish astronomer, 1473–1543. His book describing heliocentric (i.e., Sun-centered) orbits for the planets was published only a few weeks before his death. He had hesitated to release it for many years, fearing that it might be considered heretical. It is not known whether or not he saw the published book.

2. Events are described by measurements made in a coordinate system that defines a frame of reference. The question was, Where is the reference frame in which the law of inertia is valid? Newton knew that no rotating system, e.g., Earth or the Sun, would work and suggested the distant "fixed stars" as the fundamental inertial reference frame.

3. The speed of light is exactly 299,792,458 m/s. The value is set by the definition of the standard meter as being the distance light travels in 1/299,792,458 s.

4. Over time, an entire continuous spectrum of electromagnetic waves has been discovered, ranging from extremely low-frequency (radio) waves to extremely high-frequency waves (gamma rays), all moving at speed c.

5. Albert A. Michelson (1852–1931), an American experimental physicist whose development of precision optical instruments and their use in precise measurements of the speed of light and the length of the standard meter earned him the Nobel Prize in 1907. Edward W. Morley (1838–1923), an American chemist and physicist and professor at Western Reserve College when Michelson was a professor at the nearby Case School of Applied Science.

6. Albert A. Michelson and Edward W. Morley, *The American Journal of Science*, **XXXIV**, no. 203, November 1887.

7. Note that the width depends on the small angle between M_2' and M_1. A very small angle results in relatively few wide fringes, a larger angle in many narrow fringes.

8. Since the source producing the waves, the sodium lamp, was at rest relative to the interferometer, the frequency would be constant.

9. T. S. Jaseja, A. Javan, J. Murray, and C. H. Townes, *Physical Review*, **133**, A1221 (1964).

10. A. Brillet and J. Hall, *Physical Review Letters*, **42**, 549 (1979).

11. *Annalen der Physik*, **17**, 841(1905). For a translation from the original German, see the collection of original papers by Lorentz, Einstein, Minkowski, and Weyl (Dover, New York, 1923).

12. Hendrik Antoon Lorentz (1853–1928), Dutch theoretical physicist, discovered the Lorentz transformation empirically while investigating the fact that Maxwell's equations are not invariant under a Galilean transformation, although he did not recognize its importance at the time. An expert on electromagnetic theory, he was one of the first to suggest that atoms of matter might consist of charged particles whose oscillations could account for the emission of light. Lorentz used this hypothesis to explain the splitting of spectral lines in a magnetic field discovered by his student Pieter Zeeman, with whom he shared the 1902 Nobel Prize.

13. One meter of light travel time is the *time* for light to travel 1 m, i.e., $ct = 1$ m, or $t = 1$ m/3.00×10^8 m/s $= 3.3 \times 10^{-9}$ s. Similarly, 1 cm of light travel time is $ct = 1$ cm, or $t = 3.3 \times 10^{-11}$ s, and so on.

14. This example is adapted from a problem in H. Ohanian, *Modern Physics* (Englewood Cliffs, NJ: Prentice Hall, 1987).

15. Any particle that has mass.

16. Equation 1-31 would lead to imaginary values of Δs for spacelike intervals, an apparent problem. However, the geometry of spacetime is not Euclidean, but Lorentzian. While a consideration of Lorentz geometry is beyond the scope of this chapter, suffice it to say that it enables us to write $(\Delta s)^2$ for spacelike intervals as in Equation 1-33.

17. There are only two such things: photons (including those of visible light), which will be introduced in Chapter 3, and gravitons, which are the particles that transmit the gravitational force.

18. Edwin P. Hubble, *Proceedings of the National Academy of Sciences*, **15**, 168 (1929).

19. Walter Kündig, *Physical Review*, **129**, 2371 (1963).

20. C. G. Darwin, *Nature*, **180**, 976 (1957).

21. S. P. Boughn, *American Journal of Physics*, **57**, 791 (1989).

22. E. F. Taylor and J. A. Wheeler, *Spacetime Physics*, 2d ed. (New York: W. H. Freeman and Co., 1992).

23. Seen in three space dimensions by the observer in S, 50 percent of the light is concentrated in 0.06 steradian of 4π-steradian solid angle around the moving source.

24. T. Alväger and M. N. Kreisler, "Quest for Faster-Than-Light Particles," *Physical Review*, **171**, 1357 (1968).

25. Paul Ehrenfest (1880–1933), Austrian physicist and professor at the University of Leiden (the Netherlands), longtime friend and correspondent of Einstein, about whom, upon his death, Einstein wrote, "[He was] the best teacher in our profession I have ever known."

26. This experiment is described in J. C. Hafele and R. E. Keating, *Science*, **177**, 166 (1972). Although not as accurate as the experiment described in Section 1-4, its results supported the relativistic prediction.

27. R. Shaw, *American Journal of Physics*, **30**, 72 (1962).

Problems

Level I

Section 1-1 The Experimental Basis of Relativity

1-1. In episode 5 of Star Wars, the Empire's spaceships launch probe droids throughout the galaxy to seek the base of the Rebel Alliance. Suppose a spaceship moving at 2.3×10^8 m/s toward Hoth (site of the rebel base) launches a probe droid toward Hoth at 2.1×10^8 m/s relative to the spaceship. According to Galilean relativity, (*a*) What is the speed of the droid relative to Hoth? (*b*) If rebel astronomers are watching the approaching spaceship through a telescope, will they see the probe before it lands on Hoth?

1-2. In one series of measurements of the speed of light, Michelson used a path length L of 27.4 km (17 mi). (*a*) What is the time needed for light to make the round trip of distance $2L$? (*b*) What is the classical correction term in seconds in Equation 1-5, assuming Earth's speed is $v = 10^{-4}c$? (*c*) From about 1600 measurements, Michelson arrived at a result for the speed of light of 299,796 $\pm$ 4 km/s. Is this experimental value accurate enough to be sensitive to the correction term in Equation 1-5?

1-3. A shift of one fringe in the Michelson-Morley experiment would result from a difference of one wavelength or a change of one period of vibration in the round-trip travel of the light when the interferometer is rotated by 90°. What speed would Michelson have computed for Earth's motion through the ether had the experiment seen a shift of one fringe?

1-4. In the "old days" (circa 1935) pilots used to race small, relatively high-powered airplanes around courses marked by a pylon on the ground at each end of the course. Suppose two such evenly matched racers fly at airspeeds of 130 mph. (Remember, this was a long time ago!) Each flies one complete round trip of 25 miles, *but* their courses are perpendicular to each other and there is a 20-mph wind blowing steadily parallel to one course. (*a*) Which pilot wins the race and by how much? (*b*) Relative to the axes of their respective courses, what headings must the two pilots use?

1-5. Paul Ehrenfest[25] suggested the following thought experiment to illustrate the dramatically different observations that might be expected, dependent on whether light moved relative to a stationary ether or according to Einstein's second postulate:

> Suppose that you are seated at the center of a huge dark sphere with a radius of 3×10^8 m and with its inner surface highly reflective. A source at the center emits a very brief flash of light that moves outward through the darkness with uniform intensity as an expanding spherical wave.

What would you see during the first 3 seconds after the emission of the flash if (*a*) the sphere moved through the ether at a constant 30 km/s and (*b*) if Einstein's second postulate is correct?

1-6. Einstein reported that as a boy he wondered about the following puzzle. If you hold a mirror at arm's length and look at your reflection, what will happen as you begin to run? In particular, suppose you run with speed $v = 0.99c$. Will you still be able to see yourself? If so, what would your image look like, and why?

1-7. Verify by calculation that the result of the Michelson-Morley experiment places an upper limit on Earth's speed relative to the ether of about 5 km/s.

1-8. Consider two inertial reference frames. When an observer in each frame measures the following quantities, which measurements made by the two observers *must* yield the same results? Explain your reason for each answer.

(*a*) The distance between two events
(*b*) The value of the mass of a proton
(*c*) The speed of light
(*d*) The time interval between two events
(*e*) Newton's first law
(*f*) The order of the elements in the periodic table
(*g*) The value of the electron charge

Section 1-2 Einstein's Postulates

1-9. Assume that the train shown in Figure 1-14 is 1.0 km long as measured by the observer at C' and is moving at 150 km/h. What time interval between the arrival of the wave fronts at C' is measured by the observer at C in S?

1-10. Suppose that A', B', and C' are at rest in frame S', which moves with respect to S at speed v in the $+x$ direction. Let B' be located exactly midway between A' and C'. At $t' = 0$ a light flash occurs at B' and expands outward as a spherical wave. (*a*) According to an observer in S', do the wave fronts arrive at A' and C' simultaneously? (*b*) According to an observer in S, do the wave fronts arrive at A' and C' simultaneously? (*c*) If you answered no to either (*a*) or (*b*), what is the difference in their arrival times and at which point did the front arrive first?

Section 1-3 The Lorentz Transformation

1-11. Make a graph of the relativistic factor $\gamma = 1/(1 - v^2/c^2)^{1/2}$ as a function of $\beta = v/c$. Use at least 10 values of β ranging from 0 up to 0.995.

1-12. Two events happen at the same point x'_0 in frame S' at times t'_1 and t'_2. (*a*) Use Equation 1-19 to show that in frame S, the time interval between the events is greater than $t'_2 - t'_1$ by a factor γ. (*b*) Why is Equation 1-18 less convenient than Equation 1-19 for this problem?

1-13. Suppose that an event occurs in inertial frame S with coordinates $x = 75$ m, $y = 18$ m, $z = 4.0$ m at $t = 2.0 \times 10^{-5}$ s. The inertial frame S' moves in the $+x$ direction with $v = 0.85c$. The origins of S and S' coincided at $t = t' = 0$. (*a*) What are the coordinates of the event in S'? (*b*) Use the inverse transformation on the results of (*a*) to obtain the original coordinates.

1-14. Show that the null effect of the Michelson-Morley experiment can be accounted for if the interferometer arm parallel to the motion is shortened by a factor of $(1 - v^2/c^2)^{1/2}$.

1-15. Two spaceships are approaching each other. (*a*) If the speed of each is $0.9c$ relative to Earth, what is the speed of one relative to the other? (*b*) If the speed of each relative to Earth is 30,000 m/s (about 100 times the speed of sound), what is the speed of one relative to the other?

1-16. Starting with the Lorentz transformation for the components of the velocity (Equation 1-23), derive the transformation for the components of the acceleration.

1-17. Consider a clock at rest at the origin of the laboratory frame. (*a*) Draw a spacetime diagram that illustrates that this clock ticks slow when observed from the reference frame of a rocket moving with respect to the laboratory at $v = 0.8c$. (*b*) When 10 s have elapsed on the rocket clock, how many have ticked by on the lab clock?

1-18. A light beam moves along the y' axis with speed c in frame S', which is moving to the right with speed v relative to frame S. (*a*) Find u_x and u_y, the x and y components of the velocity of the light beam in frame S. (*b*) Show that the magnitude of the velocity of the light beam in S is c.

1-19. A particle moves with speed $0.9c$ along the x'' axis of frame S'', which moves with speed $0.9c$ in the positive x' direction relative to frame S'. Frame S' moves with speed $0.9c$ in the positive x direction relative to frame S. (*a*) Find the speed of the particle relative to frame S'. (*b*) Find the speed of the particle relative to frame S.

Section 1-4 Time Dilation and Length Contraction

1-20. Use the binomial expansion to derive the following results for values of $v \ll c$ and use when applicable in the problems that follow.

$$(a) \qquad \gamma \approx 1 + \frac{1}{2}\frac{v^2}{c^2}$$

$$(b) \qquad \frac{1}{\gamma} \approx 1 - \frac{1}{2}\frac{v^2}{c^2}$$

$$(c) \qquad \gamma - 1 \approx 1 - \frac{1}{\gamma} \approx \frac{1}{2}\frac{v^2}{c^2}$$

1-21. How great must the relative speed of two observers be for their time-interval measurements to differ by 1 percent (see Problem 1-20)?

1-22. A *nova* is the sudden, brief brightening of a star (see Chapter 13). Suppose Earth astronomers see two novas occur simultaneously, one in the constellation Orion (the Hunter) and the other in the constellation Lyra (the Lyre). Both nova are the same distance from Earth, $2.5 \times 10^3\ c \cdot y$, and are in exactly opposite directions from Earth. Observers on board an aircraft flying at 1000 km/h on a line from Orion toward Lyra see the same novas, but note that they are not simultaneous. (*a*) For the observers on the aircraft, how much time separates the nova? (*b*) Which one occurs first? (Assume Earth is an inertial reference frame.)

1-23. A meter stick moves parallel to its length with speed $v = 0.6c$ relative to you. (*a*) Compute the length of the stick measured by you. (*b*) How long does it take for the stick to pass you? (*c*) Draw a spacetime diagram from the viewpoint of your frame with the front of the meter stick at $x = 0$ when $t = 0$. Show how the answers to (*a*) and (*b*) are obtained from the diagram.

1-24. The proper mean lifetime of π mesons (pions) is 2.6×10^{-8} s. If a beam of such particles has speed $0.9c$, (*a*) What would their mean life be as measured in the laboratory? (*b*) How far would they travel (on the average) before they decay? (*c*) What would your answer be to part (*b*) if you neglected time dilation? (*d*) What is the interval in spacetime between creation of a typical pion and its decay?

1-25. You have been posted to a remote region of space to monitor traffic. Near the end of a quiet shift, a spacecraft streaks past. Your laser-based measuring device reports the spacecraft's length to be 85 m. The identification transponder reports it to be the NCXXB-12, a cargo craft of proper length 100 m. In transmitting your report to headquarters, what speed should you give for this spacecraft?

1-26. The light clock in the spaceship in Figure 1-25 uses a light pulse moving up the y-axis to reflect back from a mirror as the ship moves along the x-axis. Suppose instead the light pulse moves along the x'-axis between $x' = 0$ and a mirror at $x' = L$. (*a*) What is the time required for the pulse to make a round trip in the rest system of the spaceship? (*b*) What is the round-trip time in the laboratory frame? (*c*) Does the result in (*b*) agree with that expected from time dilation? Justify your answer.

1-27. Two spaceships pass each other traveling in opposite directions. A passenger on ship *A*, which she knows to be 100 m long, notes that ship *B* is moving with a speed of $0.92c$ relative to *A* and that the length of *B* is 36 m. What are the lengths of the two spaceships measured by a passenger in *B*?

1-28. A meter stick at rest in S' is tilted at an angle of 30° to the x' axis. If S' moves at $\beta = 0.8$, how long is the meter stick as measured in S and what angle does it make with the x axis?

1-29. A rectangular box at rest in S' has sides $a' = 2$ m, $b' = 2$ m, and $c' = 4$ m and is oriented as shown in Figure 1-44. S' moves with $\beta = 0.65$ with respect to the laboratory frame S. (*a*) Compute the volume of the box in S' and in S. (*b*) Draw an accurate diagram of the box as seen by an observer in S.

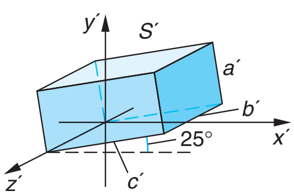


Figure 1-44

Section 1-5 The Doppler Effect

1-30. How fast must you be moving toward a red light ($\lambda = 650$ nm) for it to appear yellow ($\lambda = 590$ nm)? green ($\lambda = 525$ nm)? blue ($\lambda = 460$ nm)?

1-31. A distant galaxy is moving away from us at speed 1.85×10^7 m/s. Calculate the fractional red shift $(\lambda' - \lambda_0)/\lambda_0$ of the light from this galaxy.

1-32. The light from a nearby star is observed to be shifted toward the blue by 2 percent, i.e., $f_{obs} = 1.02 f_0$. Is the star approaching or receding from Earth? How fast is it moving? (Assume motion is directly toward or away from Earth to avoid superluminal speeds.)

1-33. Stars typically emit the red light of atomic hydrogen with wavelength 656.3 nm (called the H_α spectral line). Compute the wavelength of that light observed at Earth from stars receding directly from us with relative speed $v = 10^{-3}c$, $v = 10^{-2}c$, and $v = 10^{-1}c$.

Section 1-6 The Twin Paradox and Other Surprises

1-34. Heide boards a spaceship and travels away from Earth at a constant velocity $0.45c$ toward Betelgeuse (a red giant star in the constellation Orion). One year later on Earth clocks, Heide's twin, Hans, boards a second spaceship and follows her at a constant velocity of $0.95c$ in the same direction. (*a*) When Hans catches up to Heide, what will be the difference in their ages? (*b*) Which twin will be older?

1-35. You point a laser flashlight at the Moon, producing a spot of light on the Moon's surface. At what minimum angular speed must you sweep the laser beam in order for the light spot to streak across the Moon's surface with speed $v > c$? Why can't you transmit information between research bases on the Moon with the flying spot?

1-36. A clock is placed in a satellite that orbits Earth with a period of 108 min. (*a*) By what time interval will this clock differ from an identical clock on Earth after 1 y? (*b*) How much time will have passed on Earth when the two clocks differ by 1.0 s? (Assume special relativity applies and neglect general relativity.)

1-37. Einstein used trains for a number of relativity thought experiments since they were the fastest objects commonly recognized in those days. Let's consider a train moving at $0.65c$ along a straight track at night. Its headlight produces a beam with an angular spread of 60° according to the engineer. If you are standing alongside the track (rails are 1.5 m apart), how far from you is the train when you see its approaching headlight suddenly disappear?

Level II

1-38. In 1971 four portable atomic clocks were flown around the world in jet aircraft, two eastbound and two westbound, to test the time dilation predictions of relativity.[26] (*a*) If the westbound plane flew at an average speed of 1500 km/h relative to the surface, how long would it have had to fly for the clock on board to lose 1 second relative to the reference clock on the ground at the U.S. Naval Observatory? (*b*) In the actual experiment the planes circumflew Earth once and the observed discrepancy of the clocks was 273 ns. What was the average speed of each plane?

1-39. "Ether drag" was among the suggestions made to explain the null result of the Michelson-Morley experiment (see More section). The phenomenon of stellar aberration refutes this proposal. Suppose Earth moves relative to the ether at velocity $\boldsymbol{v}$ and a light beam (e.g., from a star) approaches Earth at an angle θ with respect to $\boldsymbol{v}$. (*a*) Show that the angle of approach in Earth's reference frame θ' is given by

$$\tan\theta' = \frac{\sin\theta}{\cos\theta + v/c}$$

(*b*) θ' is the stellar aberration angle. If $\theta = 90°$, by how much does θ' differ from 90°?

1-40. A friend of yours who is the same age as you travels to the star Alpha Centauri, which is 4 $c \cdot$y away, and returns immediately. He claims that the entire trip took just 6 years. (*a*) How fast did he travel? (*b*) How old are you when he returns? (*c*) Draw a spacetime diagram that verifies your answers to (*a*) and (*b*).

1-41. A clock is placed in a satellite that orbits Earth with a period of 90 min. By what time interval will this clock differ from an identical clock on Earth after 1 year? (Assume that special relativity applies.)

1-42. In frame S, event B occurs 2 μs after event A and at $\Delta x = 1.5$ km from event A. (*a*) How fast must an observer be moving along the $+x$ axis so that events A and B occur simultaneously? (*b*) Is it possible for event B to precede event A for some observer? (*c*) Draw a spacetime diagram that illustrates your answers to (*a*) and (*b*). (*d*) Compute the spacetime interval and proper distance between the events.

1-43. A burst of π^+ mesons travels down an evacuated beam tube at Fermilab moving at $\beta = 0.92$ with respect to the laboratory. (*a*) Compute γ for this group of pions. (*b*) The proper mean lifetime of pions is 2.6×10^{-8} s. What mean lifetime is measured in the lab? (*c*) If the burst contained 50,000 pions, how many remain after the group has traveled 50 m down the beam tube? (*d*) What would be the answer to (*c*) ignoring time dilation?

1-44. H. A. Lorentz suggested 15 years before Einstein's 1905 paper that the null effect of the Michelson-Morley experiment could be accounted for by a contraction of that arm of the interferometer lying parallel to Earth's motion through the ether to a length $L = L_p(1 - v^2/c^2)^{-1/2}$. He thought of this, incorrectly, as an actual shrinking of matter. By about how many atomic diameters would the material in the parallel arm of the interferometer have had to shrink in order to account for the absence of the expected shift of 0.4 of a fringe width? (Assume the diameter of atoms to be about 10^{-10} m.)

1-45. Observers in reference frame S see an explosion located at $x_1 = 480$ m. A second explosion occurs 5 μs later at $x_2 = 1200$ m. In reference frame S', which is moving along the $+x$ axis at speed v, the explosions occur at the same point in space. (*a*) Draw a spacetime diagram describing this situation. (*b*) Determine v from the diagram. (*c*) Calibrate the ct' axis and determine the separation in time in μs between the two explosions as measured in S'. (*d*) Verify your results by calculation.

1-46. Two spaceships, each 100 m long when measured at rest, travel toward each other with speeds of $0.85c$ relative to Earth. (*a*) How long is each ship as measured by someone on Earth? (*b*) How fast is each ship traveling as measured by an observer on the other? (*c*) How long is one ship when measured by an observer on the other? (*d*) At time $t = 0$ on Earth, the fronts of the ships are together as they just begin to pass each other. At what time on Earth are their ends together? (*e*) Sketch accurately scaled diagrams in the frame of one of the ships showing the passing of the other ship.

1-47. If v is much less than c, the Doppler frequency shift is approximately given by $\Delta f/f_0 = \pm\beta$, both classically and relativistically. A radar transmitter-receiver bounces a signal off an aircraft and observes a fractional increase in the frequency of $\Delta f/f_0 = 8 \times 10^{-7}$. What is the speed of the aircraft? (Assume the aircraft to be moving directly toward the transmitter.)

1-48. The null result of the Michelson-Morley experiment could be explained if the speed of light depended on the motion of the source relative to the observer. Consider a binary eclipsing star system, that is, a pair of stars orbiting their common center of mass with Earth lying in the orbital plane of the system, as is very nearly the case for the binary system Algol (see More section about the Michelson-Morley experiment). Assume that the stars in the system have circular orbits with a period of 115 days and that one of the stars' orbital speed is 32 km/s (about the same as Earth's orbital speed around the Sun). If the suggestion above were true, astronomers would simultaneously see two images of the star in opposition, i.e., on opposite sides of its orbit. What is the minimum distance L from Earth to the binary for this phenomenon to occur?

1-49. Frames S and S' are moving relative to each other along the x and x' axes. They set their clocks to $t = t' = 0$ when their origins coincide. In frame S, event 1 occurs at $x_1 = 1\ c \cdot \mathrm{y}$ and $t_1 = 1$ y and event 2 occurs at $x_2 = 2.0\ c \cdot \mathrm{y}$ and $t_2 = 0.5$ y. These events occur simultaneously in frame S'. (*a*) Find the magnitude and direction of the velocity of S' relative to S. (*b*) At what time do both of these events occur as measured in S'? (*c*) Compute the spacetime interval Δs between the events. (*d*) Is the interval spacelike, timelike, or lightlike? (*e*) What is the proper distance L_p between the events?

1-50. Do Problem 1-49 parts (*a*) and (*b*) using a spacetime diagram.

1-51. An observer in frame S standing at the origin observes two flashes of colored light separated spatially by $\Delta x = 2400$ m. A blue flash occurs first, followed by a red flash 5 μs later. An observer in S' moving along the x axis at speed v relative to S also observes the flashes 5 μs apart and with a separation of 2400 m, but the red flash is observed first. Find the magnitude and direction of v.

1-52. A cosmic-ray proton streaks through the lab with velocity $0.85c$ at an angle of 50° with the $+x$ direction (in the xy plane of the lab). Compute the magnitude and direction of the proton's velocity when viewed from frame S' moving with $\beta = 0.72$.

Level III

1-53. A meter stick is parallel to the x axis in S and is moving in the $+y$ direction at constant speed v_y. From the viewpoint of S show that the meter stick will appear tilted at an angle θ' with respect to the x' axis of S' moving in the $+x$ direction at $\beta = 0.65$. Compute the angle θ' measured in S'.

1-54. The equation for the spherical wave front of a light pulse that begins at the origin at time $t = 0$ is $x^2 + y^2 + z^2 - (ct)^2 = 0$. Using the Lorentz transformation, show that such a light pulse also has a spherical wave front in S' by showing that $x'^2 + y'^2 + z'^2 - (ct')^2 = 0$ in S'.

1-55. An interesting paradox has been suggested by R. Shaw[27] that goes like this. A very thin steel plate with a circular hole 1 m in diameter centered on the y axis lies parallel to the xz plane

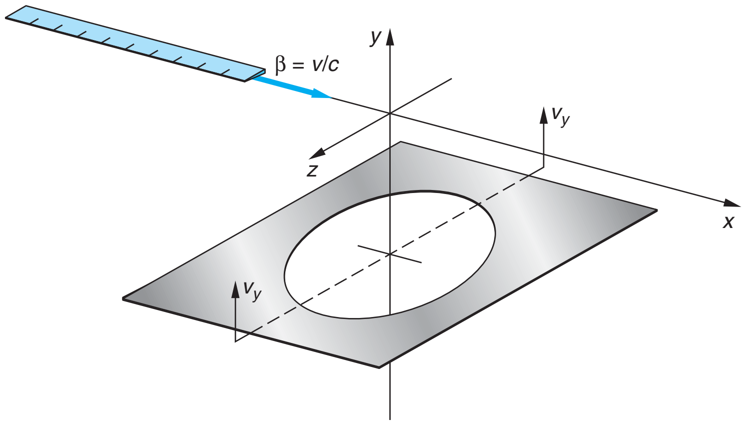


Figure 1-45

in frame S and moves in the $+y$ direction at constant speed v_y as illustrated in Figure 1-45. A meter stick lying on the x axis moves in the $+x$ direction with $\beta = v/c$. The steel plate arrives at the $y = 0$ plane at the same instant that the center of the meter stick reaches the origin of S. Since the meter stick is observed by observers in S to be contracted, it passes through the 1-m hole in the plate with no problem. A paradox appears to arise when one considers that an observer in S', the rest system of the meter stick, measures the diameter of the hole in the plate to be contracted in the x dimension and, hence, becomes too small to pass the meter stick, resulting in a collision. Resolve the paradox. Will there be a collision?

1-56. Two events in S are separated by a distance $D = x_2 - x_1$ and a time $T = t_2 - t_1$. (*a*) Use the Lorentz transformation to show that in frame S', which is moving with speed v relative to S, the time separation is $t_2 - t_1 = \gamma(T - vD/c^2)$. (*b*) Show that the events can be simultaneous in frame S' only if D is greater than cT. (*c*) If one of the events is the *cause* of the other, the separation D must be less than cT since D/c is the smallest time that a signal can take to travel from x_1 to x_2 in frame S. Show that if D is less than cT, t_2' is greater than t_1' in all reference frames. (*d*) Suppose that a signal could be sent with speed $c' > c$ so that in frame S the cause precedes the effect by the time $T = D/c'$. Show that there is then a reference frame moving with speed v less than c in which the effect precedes the cause.

1-57. Two observers agree to test time dilation. They use identical clocks and one observer in frame S' moves with speed $v = 0.6c$ relative to the other observer in frame S. When their origins coincide, they start their clocks. They agree to send a signal when their clocks read 60 min and to send a confirmation signal when each receives the other's signal. (*a*) When does the observer in S receive the first signal from the observer in S'. (*b*) When does he receive the confirmation signal? (*c*) Make a table showing the times in S when the observer sent the first signal, received the first signal, and received the confirmation signal. How does this table compare with one constructed by the observer in S'?

1-58. The compact disk in a CD-ROM drive rotates with angular speed ω. There is a clock at the center of the disk and one at a distance r from the center. In an inertial reference frame, the clock at distance r is moving with speed $u = r\omega$. Show that from time dilation in special relativity, time intervals Δt_o for the clock at rest and Δt_r for the moving clock are related by

$$\frac{\Delta t_r - \Delta t_o}{\Delta t_o} \approx \frac{r^2\omega^2}{2c^2} \qquad \text{if } r\omega \ll c$$

1-59. Two rockets, A and B, leave a space station with velocity vectors $\mathbf{v}_A$ and $\mathbf{v}_B$ relative to the station frame S, perpendicular to each other. (*a*) Determine the velocity of A relative to B, $\mathbf{v}_{BA}$. (*b*) Determine the velocity of B relative to A, $\mathbf{v}_{AB}$. (*c*) Explain why $\mathbf{v}_{AB}$ and $\mathbf{v}_{BA}$ do not point in opposite directions.

1-60. Suppose a system S consisting of a cubic lattice of meter sticks and synchronized clocks, e.g., the eight clocks closest to you in Figure 1-13, moves from left to right (the $+x$ direction) at high speed. The meter sticks parallel to the x direction are, of course, contracted and the cube would be *measured* by an observer in a system S' to be foreshortened in that direction. However, recalling that your eye constructs images from light waves that reach it simultaneously, not those leaving the source simultaneously, sketch what your eye would *see* in this case. Scale contractions and show any angles accurately. (Assume the moving cube to be farther than 10 m from your eye.)

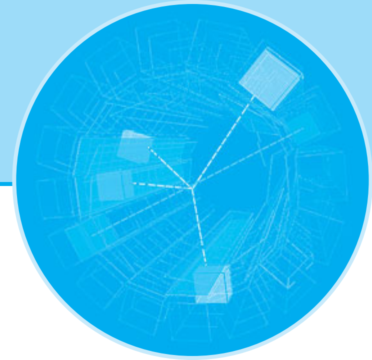
1-61. Figure 1-11*b* (in the More section about the Michelson-Morley experiment) shows an eclipsing binary. Suppose the period of the motion is T and the binary is a distance L from Earth, where L is sufficiently large so that points A and B in Figure 1-11*b* are a half orbit apart. Consider the motion of one of the stars and (*a*) show that the star would appear to move from A to B in time $T/2 + 2Lv/(c^2 - v^2)$ and from B to A in time $T/2 - 2Lv/(c^2 - v^2)$, assuming classical velocity addition applies to light, i.e., that emission theories of light were correct. (*b*) What rotational period would cause the star to appear to be at both A and B simultaneously?

1-62. Show that if a particle moves at an angle θ with respect to the x axis with speed u in system S, it moves at an angle θ' with the x' axis in S' given by

$$\tan\theta' = \frac{\sin\theta}{\gamma(\cos\theta - v/u)}$$

1-63. Like jets emitted from some galaxies (see Figure 1-41), some distant astronomical objects can appear to travel at speeds greater than c across our line of sight. Suppose distant galaxy AB15 moving with velocity v at an angle θ with respect to the direction toward Earth emits two bright flashes of light separated by time Δt on the galaxy AB15 local clock. Show that (*a*) the time interval $\Delta t_{\text{Earth}} = \Delta t(1 - \beta\cos\theta)$ and (*b*) the apparent speed of AB15 measured by observers on Earth is $v_{\text{app}} = \dfrac{\Delta x_{\text{Earth}}}{\Delta t_{\text{Earth}}} = \dfrac{\beta\sin\theta}{1 - \beta\cos\theta}$. (*c*) For $\beta = 0.75$, compute the value of θ for which $v_{\text{app}} = c$.

CHAPTER 2



Relativity II

In the opening section of Chapter 1, we discussed the classical observation that, if Newton's second law, $\mathbf{F} = m\mathbf{a}$, holds in a particular reference frame, it also holds in any other reference frame that moves with constant velocity relative to it, i.e., in any inertial frame. As shown in Section 1-1, the Galilean transformation (Equation 1-2) leads to the same accelerations $a'_x = a_x$ in both frames, and forces such as those due to stretched springs are also the same in both frames. However, according to the Lorentz transformation, accelerations are not the same in two such reference frames. If a particle has acceleration a_x and velocity u_x in frame S, its acceleration in S', obtained by computing du'_x/dt' from Equation 1-22, is

$$a'_x = \frac{a_x}{\gamma^3(1 - vu_x/c^2)^3} \qquad \textbf{2-1}$$

Thus, F/m must transform in a similar way, or else Newton's second law, $\mathbf{F} = m\mathbf{a}$, does not hold.

It is reasonable to expect that $\mathbf{F} = m\mathbf{a}$ does *not* hold at high speeds, for this equation implies that a constant force will accelerate a particle to unlimited velocity if it acts for a long time. However, if a particle's velocity were greater than c in some reference frame S, we could not transform from S to the rest frame of the particle because γ becomes imaginary when $v > c$. We can show from the velocity transformation that, if a particle's velocity is less than c in some frame S, it is less than c in all frames moving relative to S with $v < c$. This result leads us to expect that particles never have speeds greater than c. Thus, we expect that Newton's second law $\mathbf{F} = m\mathbf{a}$ is not relativistically invariant. We will, therefore, need a new law of motion, but one that reduces to Newton's classical version when $\beta(=v/c) \rightarrow 0$, since $\mathbf{F} = m\mathbf{a}$ is consistent with experimental observations when $\beta \ll 1$.

In this chapter we will explore the changes in classical dynamics that are dictated by relativity theory, directing particular attention to the same concepts around which classical mechanics was developed, namely, mass, momentum, and energy. We will find these changes to be every bit as dramatic as those we encountered in Chapter 1, including a Lorentz transformation for momentum and energy and a new invariant quantity to stand beside the invariant spacetime interval Δs. Then, in the latter part of the chapter, we will briefly turn our attention to noninertial, or accelerated, reference frames, the realm of the theory of general relativity.

2-1 Relativistic Momentum

Among the most powerful fundamental concepts that you have studied in physics until now are the ideas of conservation of momentum and conservation of total energy. As we will discuss a bit further in Chapter 12, each of these fundamental laws arises because of a particular symmetry that exists in the laws of physics. For example, the conservation of total energy in classical physics is a consequence of the symmetry, or invariance, of the laws of physics to translations in time. As a consequence, Newton's laws work exactly the same way today as they did when he first wrote them down. The conservation of momentum arises from the invariance of physical laws to translations in space. Indeed, Einstein's first postulate and the resulting Lorentz transformation (Equations 1-18 and 1-19) guarantee this latter invariance in all inertial frames.

The simplicity and universality of these conservation laws leads us to seek equations for relativistic mechanics, replacing Equation 1-1 and others, that are consistent with momentum and energy conservation and are also invariant under a Lorentz transformation. However, it is straightforward to show that the momentum, as formulated in classical mechanics, does not result in relativistic invariance of the law of conservation of momentum. To see that this is so, we will look at an isolated collision between two masses, where we avoid the question of how to transform forces because the net external force is zero. In classical mechanics, the total momentum $\mathbf{p} = \Sigma m_i \mathbf{u}_i$ is conserved. We will see that relativistically, conservation of the quantity $\Sigma m_i \mathbf{u}_i$ is an approximation that holds only at low speeds.

Consider one observer in frame S with a ball A and another in S' with ball B. The balls each have mass m and are identical when measured at rest. Each observer throws his ball along his y axis with speed u_0 (measured in his own frame) so that the balls collide.[1] Assuming the balls to be perfectly elastic, each observer will see his ball rebound with its original speed u_0. If the total momentum is to be conserved, the y component must be zero because the momentum of each ball is merely reversed by the collision. However, if we consider the relativistic velocity transformation, we can see that the quantity mu_y does not have the same magnitude for each ball as seen by either observer.

Let us consider the collision as seen in frame S (Figure 2-1a). In this frame ball A moves along the y axis with velocity $u_{yA} = u_0$. Ball B has x component of velocity $u_{xB} = v$ and y component

$$u_{yB} = u'_{yB}/\gamma = -u_0\sqrt{1 - v^2/c^2} \tag{2-2}$$

Here we have used the velocity transformation (Equation 1-22) and the facts that u'_{yB} is just $-u_0$ and $u'_{xB} = 0$. We see that the y component of the velocity of ball B is smaller in magnitude than that of ball A. The quantity $(1 - v^2/c^2)^{1/2}$ comes from the time dilation factor. The time taken for ball B to travel a given distance along the y axis in S is greater than the time measured in S' for the ball to travel this same distance.

Thus, in S the total y component of classical momentum is not zero. Since the y components of the velocities are reversed in an elastic collision, momentum as defined by $\mathbf{p} = \Sigma m\mathbf{u}$ is not conserved in S. Analysis of this problem in S' leads to the same conclusion (Figure 2-1b), since the roles of A and B are simply interchanged.[2] In the classical limit $v \ll c$, momentum is conserved, of course, because in that limit $\gamma \approx 1$ and $u_{yB} \approx u_0$.

The reason for defining momentum as $\Sigma m\mathbf{u}$ in classical mechanics is that this quantity is conserved when there is no external force, as in our collision example. We now see that this quantity is conserved only in the approximation $v \ll c$. We will define the *relativistic momentum* $\mathbf{p}$ of a particle to have the following properties:

1. $\mathbf{p}$ is conserved in collisions.
2. $\mathbf{p}$ approaches $m\mathbf{u}$ as u/c approaches zero.

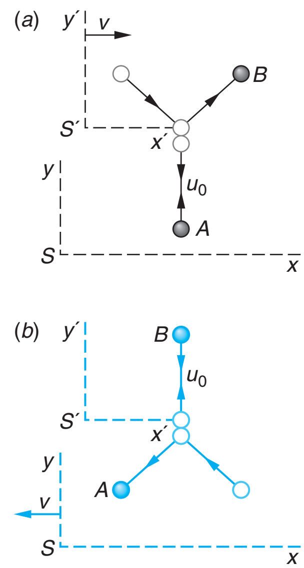


Figure 2-1 (a) Elastic collision of two identical balls as seen in frame S. The vertical component of the velocity of ball B is u_0/γ in S if it is u_0 in S'. (b) The same collision as seen in S'. In this frame, ball A has vertical component of velocity u_0/γ.

Let's apply the first of these conditions to the collision of the two balls that we just discussed, noting two important points. First, for each observer in Figure 2-1, the speed of each ball is unchanged by the elastic collision. It is either u_0 (for the observer's own ball) or $(u_y^2 + v^2)^{1/2} = u$ (for the other ball). Second, the failure of the conservation of momentum in the collision we described can't be due to the velocities because we used the Lorentz transformation to find the y components. It must have something to do with the mass! Let us write down the conservation of the y component of the momentum *as observed in S*, keeping the masses of the two balls straight by writing $m(u_0)$ for the S observer's own ball and $m(u)$ for the S' observer's ball.

$$\underset{\text{(before collision)}}{m(u_0)u_0 - m(u)u_{yB}} = \underset{\text{(after collision)}}{-m(u_0)u_0 + m(u)u_{yB}} \tag{2-3}$$

Equation 2-3 can be readily rewritten as

$$\frac{m(u)}{m(u_0)} = \frac{u_0}{u_{yB}} \tag{2-4}$$

If u_0 is small compared to the relative speed v of the reference frames, then it follows from Equation 2-2 that $u_{yB} \ll v$ and, therefore, $u \approx v$.

If we can now imagine the limiting case where $u_0 \rightarrow 0$, i.e., where each ball is at rest in its "home" frame so that the collision becomes a "grazing" one as B moves past A at speed $v = u$, then we conclude from Equations 2-2 and 2-4 that in order for Equation 2-3 to hold, i.e., for the momentum to be conserved,

$$\frac{m(u = v)}{m(u_0 = 0)} = \frac{u_0}{u_0\sqrt{1 - v^2/c^2}}$$

or

$$m(u) = \frac{m}{\sqrt{1 - v^2/c^2}} \tag{2-5}$$

Equation 2-5 says that the observer in S *measures* the mass of ball B, moving relative to him at speed u, as equal to $1/(1 - v^2/c^2)^{1/2}$ times the rest mass of the ball, or its mass measured in the frame in which it is at rest. Notice that observers always measure the mass of an object that is in motion with respect to them to be larger than the value measured when the object is at rest.

The design and construction of large particle accelerators throughout the world, such as CERN's LHC, are based directly on the relativistic expressions for momentum and energy.

Thus, we see that the law of conservation of momentum will be valid in relativity, provided that we write the momentum **p** of an object with rest mass m moving with velocity **u** relative to an inertial system S to be

$$\mathbf{p} = \frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}} \tag{2-6}$$

where u is the speed of the particle. We therefore take this equation as the definition of relativistic momentum. It is clear that this definition meets our second criterion because the denominator approaches 1 when u is much less than c. From this definition, the momenta of the two balls A and B in Figure 2-1 as seen in S are

$$p_{yA} = \frac{mu_0}{\sqrt{1 - u_0^2/c^2}} \qquad p_{yB} = \frac{mu_{yB}}{\sqrt{1 - (u_{xB}^2 + u_{yB}^2)/c^2}}$$

where $u_{yB} = u_0(1 - v^2/c^2)^{1/2}$ and $u_{xB} = v$. It is similarly straightforward to show that $p_{yB} = -p_{yA}$. Because of the similarity of the factor $1/\sqrt{1 - u^2/c^2}$ and γ in the Lorentz transformation, Equation 2-6 is often written

$$\mathbf{p} = \gamma m\mathbf{u} \quad \text{with} \quad \gamma = \frac{1}{\sqrt{1 - u^2/c^2}} \tag{2-7}$$

This use of the symbol γ for two different quantities causes some confusion; the notation is standard, however, and simplifies many of the equations. We will use this notation except when we are also considering transformations between reference frames. Then, to avoid confusion, we will write out the factor $1/(1 - u^2/c^2)^{1/2}$ and reserve γ for $1/(1 - v^2/c^2)^{1/2}$, where v is the relative speed of the frames. Figure 2-2 shows a graph of the magnitude of **p** as a function of u/c. The quantity $m(u)$ in Equation 2-5 is sometimes called the *relativistic mass*; however, we will avoid using the term or a symbol for relativistic mass: in this book m *always* refers to the mass

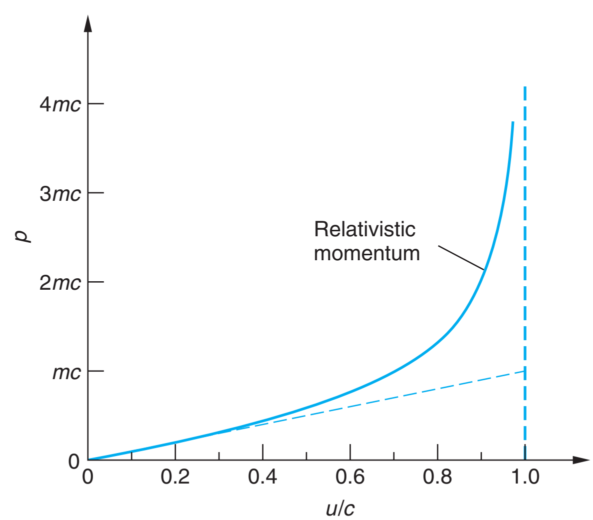


Figure 2-2 Relativistic momentum as given by Equation 2-6 versus u/c, where u = speed of the object relative to an observer. The magnitude of the momentum p is plotted in units of mc. The fainter dashed line shows the classical momentum mu for comparison.

measured in the rest frame. In this we are following Einstein's view. In a letter to a colleague in 1948 he wrote:[3]

> It is not good to introduce the concept of mass $M = m/(1 - v^2/c^2)^{1/2}$ of a body for which no clear definition can be given. It is better to introduce no other mass than "the rest mass" m. Instead of introducing M, it is better to mention the expression for the momentum and energy of a body in motion.

EXAMPLE 2-1 Measured Values of Moving Mass For what value of u/c will the measured mass of an object γm exceed the rest mass by a given fraction f?

SOLUTION

From Equation 2-5 we see that

$$f = \frac{\gamma m - m}{m} = \gamma - 1 = \frac{1}{\sqrt{1 - u^2/c^2}} - 1$$

Solving for u/c,

$$1 - u^2/c^2 = \frac{1}{(f + 1)^2} \longrightarrow u^2/c^2 = 1 - \frac{1}{(f + 1)^2}$$

or

$$u/c = \frac{\sqrt{f(f + 2)}}{f + 1}$$

from which we can compute the table of values below or the value of u/c for any other f. Note that the value of u/c that results in a given fractional increase f in the measured value of the mass is independent of m. A diesel locomotive moving at a particular u/c will be observed to have the same f as a proton moving with that u/c.

f	u/c	Example
10^{-12}	1.4×10^{-6}	jet fighter aircraft
5×10^{-9}	0.0001	Earth's orbital speed
0.0001	0.014	50-eV electron
0.01 (1%)	0.14	quasar 3C273
1.0 (100%)	0.87	quasar 0Q172
10	0.996	muons from cosmic rays
100	0.99995	some cosmic ray protons

EXAMPLE 2-2 Momentum of a Rocket A high-speed interplanetary probe with a mass $m = 50{,}000$ kg has been sent toward Pluto at a speed $u = 0.8c$. What is its momentum as measured by Mission Control on Earth? If, preparatory to landing on Pluto, the probe's speed is reduced to $0.4c$, by how much does its momentum change?

SOLUTION

1. Assuming that the probe travels in a straight line toward Pluto, its momentum along that direction is given by Equation 2-6:

$$p = \frac{mu}{\sqrt{1 - u^2/c^2}} = \frac{(50{,}000 \text{ kg})(0.8c)}{\sqrt{1 - (0.8c)^2/c^2}}$$
$$= 6.7 \times 10^4 \, c \cdot \text{kg} = 2.0 \times 10^{13} \text{ kg} \cdot \text{m/s}$$

2. When the probe's speed is reduced, the momentum declines along the relativistic momentum curve in Figure 2-2. The new value is computed from the ratio:

$$\frac{p_{0.4c}}{p_{0.8c}} = \frac{m(0.4c)/\sqrt{1 - (0.4)^2}}{m(0.8c)/\sqrt{1 - (0.8)^2}}$$
$$= \frac{1}{2} \frac{\sqrt{1 - (0.8)^2}}{\sqrt{1 - (0.4)^2}}$$
$$= 0.33$$

3. The reduced momentum $p_{0.4c}$ is then given by

$$p_{0.4c} = 0.33 p_{0.8c}$$
$$= (0.33)(6.7 \times 10^4 \, c \cdot \text{kg})$$
$$= 2.2 \times 10^4 \, c \cdot \text{kg}$$
$$= 6.6 \times 10^{12} \text{ kg} \cdot \text{m/s}$$

Remarks: *Notice from Figure 2-2 that the incorrect classical value of $p_{0.8c}$ would have been 4.0×10^4 $c \cdot$ kg. Also, while the probe's speed was decreased to 1/2 its initial value, the momentum was decreased to 1/3 of the initial value.*

Question

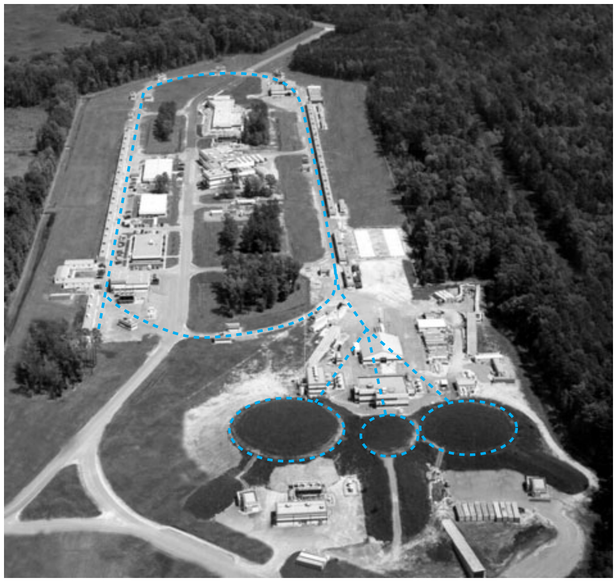
1. In our discussion of the inelastic collision of balls A and B, the collision was a "grazing" one in the limiting case. Suppose instead that the collision is a "head-on" one along the x axis. If the speed of S' (i.e., ball B) is low, say, $v = 0.1c$, what would a spacetime diagram of the collision look like?

2-2 Relativistic Energy

As noted in the preceding section, the fundamental character of the principle of conservation of total energy leads us to seek a definition of total energy in relativity that preserves the invariance of that conservation law in transformations between inertial systems. As with the definition of the relativistic momentum, Equation 2-6, we will require that the *relativistic total energy* E satisfy two conditions:

1. The total energy E of any isolated system is conserved.
2. E will approach the classical value when u/c approaches zero.

Let us first find a form for E that satisfies the second condition and then see if it also satisfies the first. We have seen that the quantity $m\mathbf{u}$ is not conserved in collisions but that $\gamma m\mathbf{u}$ is, with $\gamma = 1/(1 - u^2/c^2)^{1/2}$. We have also noted that Newton's second



Aerial view of the Jefferson Laboratory's Continuous Electron Beam Accelerator Facility (CEBAF) in Virginia. The dashed line indicates the location of the underground accelerator, where electrons are accelerated to 6 GeV and reach speeds of more than 99.99 percent of the speed of light. The circles outline the experiment halls, also underground. [*Thomas Jefferson National Accelerator Facility/U.S. Department of Energy.*]

law in the form $\mathbf{F} = m\mathbf{a}$ cannot be correct relativistically, one reason being that it leads to the conservation of $m\mathbf{u}$. We can get a hint of the relativistically correct form of the second law by writing it $\mathbf{F} = d\mathbf{p}/dt$. This equation is relativistically correct *if* the relativistic momentum $\mathbf{p}$ is used. We thus define the force in relativity to be

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d(\gamma m\mathbf{u})}{dt} \tag{2-8}$$

Now, as in classical mechanics, we will define kinetic energy E_k as the work done by a net force in accelerating a particle from rest to some velocity u. Considering motion in one dimension only, we have

$$E_k = \int_{u=0}^{u} F\,dx = \int_0^u \frac{d(\gamma mu)}{dt}\,dx = \int_0^u u\,d(\gamma mu)$$

using $u = dx/dt$. The computation of the integral in this equation is not difficult but requires a bit of algebra. It is left as an exercise (Problem 2-2) to show that

$$d(\gamma mu) = m\left(1 - \frac{u^2}{c^2}\right)^{-3/2} du$$

Substituting this into the integrand of the equation for E_k above, we obtain

$$\begin{aligned} E_k &= \int_0^u u\,d(\gamma mu) = \int_0^u m\left(1 - \frac{u^2}{c^2}\right)^{-3/2} u\,du \\ &= mc^2\left(\frac{1}{\sqrt{1 - u^2/c^2}} - 1\right) \end{aligned}$$

or

$$E_k = \gamma mc^2 - mc^2 \tag{2-9}$$

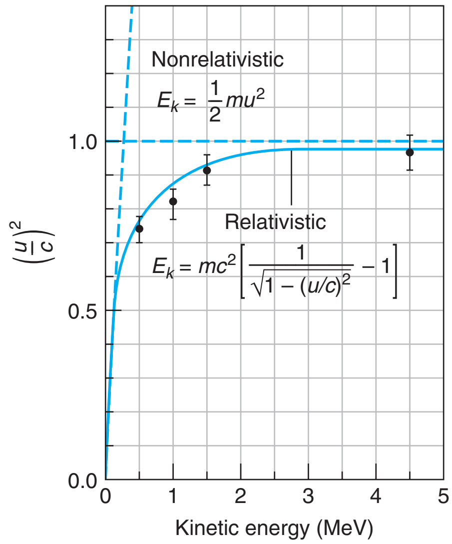


Figure 2-3 Experimental confirmation of the relativistic relation for kinetic energy. Electrons were accelerated to energies up to several MeV in large electric fields, and their velocities were determined by measuring their time of flight over 8.4 m. Note that when the velocity $u \ll c$, the relativistic and nonrelativistic (i.e., classical) relations are indistinguishable. [*W. Bertozzi, American Journal of Physics,* ***32****, 551 (1964).*]

Equation 2-9 defines the *relativistic kinetic energy*. Notice that, as we warned earlier, E_k is *not* $mu^2/2$ or even $\gamma mu^2/2$. This is strikingly evident in Figure 2-3. However, consistent with our second condition on the relativistic total energy E, Equation 2-9 does approach $mu^2/2$ when $u \ll c$. We can check this assertion by noting that for $u/c \ll 1$, expanding γ by the binomial theorem yields

$$\gamma = \left(1 - \frac{u^2}{c^2}\right)^{-1/2} = 1 + \frac{1}{2}\frac{u^2}{c^2} + \cdots$$

and thus

$$E_k = mc^2\left(1 + \frac{1}{2}\frac{u^2}{c^2} + \cdots - 1\right) \approx \frac{1}{2}mu^2$$

The expression for kinetic energy in Equation 2-9 consists of two terms. One term, γmc^2, depends on the speed of the particle (through the factor γ), and the other term, mc^2, is independent of the speed. The quantity mc^2 is called the *rest energy* of the particle, i.e., the energy associated with the rest mass m. The relativistic total energy E is then defined as the sum of the kinetic energy and the rest energy.

$$E = E_k + mc^2 = \gamma mc^2 = \frac{mc^2}{\sqrt{1 - u^2/c^2}} \qquad \textbf{2-10}$$

Thus, the work done by a net force increases the energy of the system from the rest energy mc^2 to γmc^2 (or increases the measured mass from m to γm).

For a particle at rest relative to an observer, $E_k = 0$, and Equation 2-10 becomes perhaps the most widely recognized equation in all of physics, Einstein's famous $E = mc^2$. When $u \ll c$, Equation 2-10 can be written as

$$E = \frac{1}{2}mu^2 + mc^2$$

Before the development of relativity theory, it was thought that mass was a conserved quantity;[4] consequently, m would always be the same before and after an interaction or event and mc^2 would therefore be constant. Since the zero of energy is arbitrary, we are always free to include an additive constant; therefore, our definition of the relativistic total energy reduces to the classical kinetic energy for $u \ll c$ and our second condition on E is satisfied.[5]

Be very careful to understand Equation 2-10 correctly. It defines the total energy E, and E is what we are seeking to conserve for isolated systems in all inertial frames, *not* E_k and *not* mc^2. Remember, too, the distinction between *conserved* quantities and *invariant* quantities. The former have the same value before and after an interaction in a particular reference frame. The latter have the same value when measured by observers in different reference frames. Thus, we are not requiring observers in relatively moving inertial frames to measure the same values for E, but rather that E be unchanged in interactions as measured in each frame. To assist us in showing that E as defined by Equation 2-10 is conserved in relativity, we will first see how E and **p** transform between inertial reference frames.

Lorentz Transformation of E and $\mathbf{p}$

Consider a particle of rest mass m that has an arbitrary velocity **u** with respect to frame S, as shown in Figure 2-4. System S' is a second inertial frame moving in the $+x$ direction. The particle's momentum and energy are given in the S and S' systems, respectively, by,

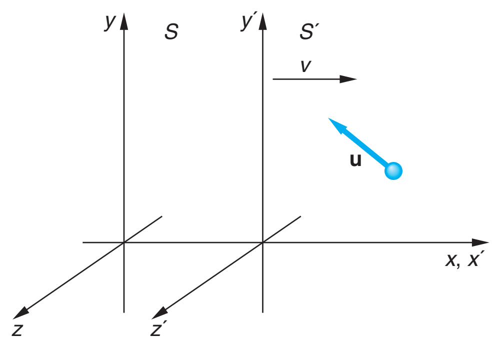


Figure 2-4 Particle of mass m moves with velocity **u** measured in S. System S' moves in the $+x$ direction at speed v. The Lorentz velocity transformation makes possible determination of the relations connecting measurements of the total energy and the components of the momentum in the two frames of reference.

In S:

$$\begin{aligned} E &= \gamma mc^2 \\ p_x &= \gamma m u_x \\ p_y &= \gamma m u_y \\ p_z &= \gamma m u_z \end{aligned} \qquad \textbf{2-11}$$

where

$$\gamma = 1/\sqrt{1 - u^2/c^2}$$

In S':

$$\begin{aligned} E' &= \gamma' mc^2 \\ p'_x &= \gamma' m u'_x \\ p'_y &= \gamma' m u'_y \\ p'_z &= \gamma' m u'_z \end{aligned} \qquad \textbf{2-12}$$

where

$$\gamma' = 1/\sqrt{1 - u'^2/c^2}$$

Developing the Lorentz transformation for E and **p** requires that we first express γ' in terms of quantities measured in S. (We could just as well express γ in terms of primed quantities. Since this is relativity, it makes no difference which we choose.) The result is

$$\frac{1}{\sqrt{1 - u'^2/c^2}} = \gamma \frac{(1 - vu_x/c^2)}{\sqrt{1 - u^2/c^2}} \quad \text{where now} \quad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \qquad \textbf{2-13}$$

Substituting Equation 2-13 into the expression for E' in Equation 2-12 yields

$$E' = \frac{mc^2}{\sqrt{1 - u'^2/c^2}} = \gamma\left[\frac{mc^2}{\sqrt{1 - u^2/c^2}} - \frac{mc^2 vu_x/c^2}{\sqrt{1 - u^2/c^2}}\right]$$

The first term in the brackets you will recognize as E, and the second term, canceling the c^2 factors, as vp_x from Equation 2-11. Thus, we have

$$E' = \gamma(E - vp_x) \tag{2-14}$$

Similarly, substituting Equation 2-13 and the velocity transformation for u'_x into the expression for p'_x in Equations 2-12 yields

$$p'_x = \frac{mu'_x}{\sqrt{1 - u'^2/c^2}} = \gamma\left[\frac{mu_x}{\sqrt{1 - u^2/c^2}} - \frac{mv}{\sqrt{1 - u^2/c^2}}\right]$$

The first term in the brackets is p_x from Equation 2-11, and, because $m(1 - u^2/c^2)^{-1/2} = E/c^2$, the second term is vE/c^2. Thus, we have

$$p'_x = \gamma(p_x - vE/c^2) \tag{2-15}$$

Using the same approach, we can show (Problem 2-46) that

$$p'_y = p_y \quad \text{and} \quad p'_z = p_z$$

Together these relations are the *Lorentz transformation for momentum and energy*:

$$\begin{aligned} p'_x &= \gamma(p_x - vE/c^2) \qquad & p'_y &= p_y \\ E' &= \gamma(E - vp_x) \qquad & p'_z &= p_z \end{aligned} \tag{2-16}$$

The inverse transformation is

$$\begin{aligned} p_x &= \gamma(p'_x + vE'/c^2) \qquad & p_y &= p'_y \\ E &= \gamma(E' + vp'_x) \qquad & p_z &= p'_z \end{aligned} \tag{2-17}$$

with

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \frac{1}{\sqrt{1 - \beta^2}}$$

Note the striking similarity between Equations 2-16 and 2-17 and the Lorentz transformation of the space and time coordinates, Equations 1-18 and 1-19. The momentum $\mathbf{p}(p_x, p_y, p_z)$ transforms in relativity exactly like $\mathbf{r}(x, y, z)$, and the total energy E transforms like the time t. We will return to this remarkable result and related matters shortly, but first let's do some examples and then, as promised, show that the energy as defined by Equation 2-10 is conserved in relativity.

EXAMPLE 2-3 Transforming Energy and Momentum Suppose a micrometeorite of mass 10^{-9} kg moves past Earth at a speed of $0.01c$. What values will be measured for the energy and momentum of the particle by an observer in a system S' moving relative to Earth at $0.5c$ in the same direction as the micrometeorite?

SOLUTION

Taking the direction of the micrometeorite's travel to be the x axis, the energy and momentum as measured by the Earth observer are, using the $u \ll c$ approximation of Equation 2-10:

$$E \approx \frac{1}{2}mu^2 + mc^2 = 10^{-9}\,\text{kg}[(0.01c)^2/2 + c^2]$$

$$E \approx 1.00005 \times 10^{-9}\,c^2\ \text{J}$$

and

$$p_x = mu_x = (10^{-9}\,\text{kg})(0.01c) = 10^{-11}c\ \text{kg}\cdot\text{m/s}$$

For this situation $\gamma = 1.1547$, so in S' the measured values of the energy and momentum will be

$$E' = \gamma(E - vp_x) = (1.1547)[1.00005 \times 10^{-9}c^2 - (0.5c)(10^{-11}c)]$$

$$E' = (1.1547)(1.00005 \times 10^{-9} - 0.5 \times 10^{-11})c^2$$

$$E' = 1.14898 \times 10^{-9}\,c^2\ \text{J}$$

and

$$p'_x = \gamma(p_x - vE/c^2) = (1.1547)[10^{-11}c - (0.5c)(1.00005 \times 10^{-9}c^2)/c^2]$$

$$p'_x = (1.1547)(10^{-11} - 5.00025 \times 10^{-10})c$$

$$p'_x = -5.66 \times 10^{-10}\,c\ \text{kg}\cdot\text{m/s} = -56.6 \times 10^{-11}\,c\ \text{kg}\cdot\text{m/s}$$

Thus, the observer in S' measures a total energy nearly 15 percent larger and a momentum more than 50 times greater and in the $-x$ direction.

EXAMPLE 2-4 A More Difficult Lorentz Transformation of Energy Suppose that a particle with mass m and energy E is moving toward the origin of a system S such that its velocity $\mathbf{u}$ makes an angle α with the y axis, as shown in Figure 2-5. Using the Lorentz transformation for energy and momentum, determine the energy E' of the particle measured by an observer in S', which moves relative to S so that the particle moves along the y' axis.

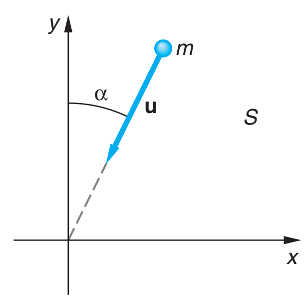


Figure 2-5 The system discussed in Example 2-4.

SOLUTION

System S' moves in the $-x$ direction at speed $u \sin\alpha$, as determined from the Lorentz velocity transformation for $u'_x = 0$. Thus, $v = -u \sin\alpha$. Also,

$$E = mc^2/\sqrt{1 - u^2/c^2} \qquad p = mu/\sqrt{1 - u^2/c^2}$$

and from the latter,

$$p = -\left(mu/\sqrt{1 - u^2/c^2}\right)\sin\alpha$$

In S' the energy will be

$$\begin{aligned} E' &= \gamma(E - vp_x) \\ &= \frac{1}{\sqrt{1 - v^2/c^2}}[E - (u\sin\alpha)\left(m - \mu/\sqrt{1 - v^2/c^2}\right)\sin\alpha] \\ &= \frac{1}{\sqrt{1 - u^2\sin^2\alpha/c^2}}[E - \left(m/\sqrt{1 - u^2/c^2}\right)u^2\sin^2\alpha] \end{aligned}$$

Multiplying the second term in the brackets by c^2/c^2 and factoring an E from both terms yield

$$E' = E\sqrt{1 - (u^2/c^2)\sin\alpha}$$

Since $u < c$ and $\sin^2\alpha \leq 1$, we see that $E' < E$, except for $\alpha = 0$ when $E' = E$, in which case S and S' are the same system. Note, too, that for $\alpha > 0$, if $u \rightarrow c$, $E' \rightarrow E\cos\alpha$. As we will see later, this is the case for light.

Question

2. Recalling the results of the measurements of time and space intervals by observers in motion relative to clocks and measuring rods, discuss the results of corresponding measurements of energy and momentum changes.

Conservation of Energy

As with our discussion of momentum conservation in relativity, let us consider a collision of two identical particles, each with rest mass m. This time, for a little variety, we will let the collision be completely inelastic—i.e., when the particles collide, they stick together. In the system S', called the *zero momentum frame*, the particles approach each other along the x' axis with equal speeds u—hence equal and opposite momenta—as illustrated in Figure 2-6*a*. In this frame the collision results in the formation of a composite particle of mass M at rest in S'. If S' moves with respect to a second frame S at speed $v = u$ in the x direction, then the particle on the right before the collision will be at rest in S and the composite particle will move to the right at speed u in that frame. This situation is illustrated in Figure 2-6*b*.

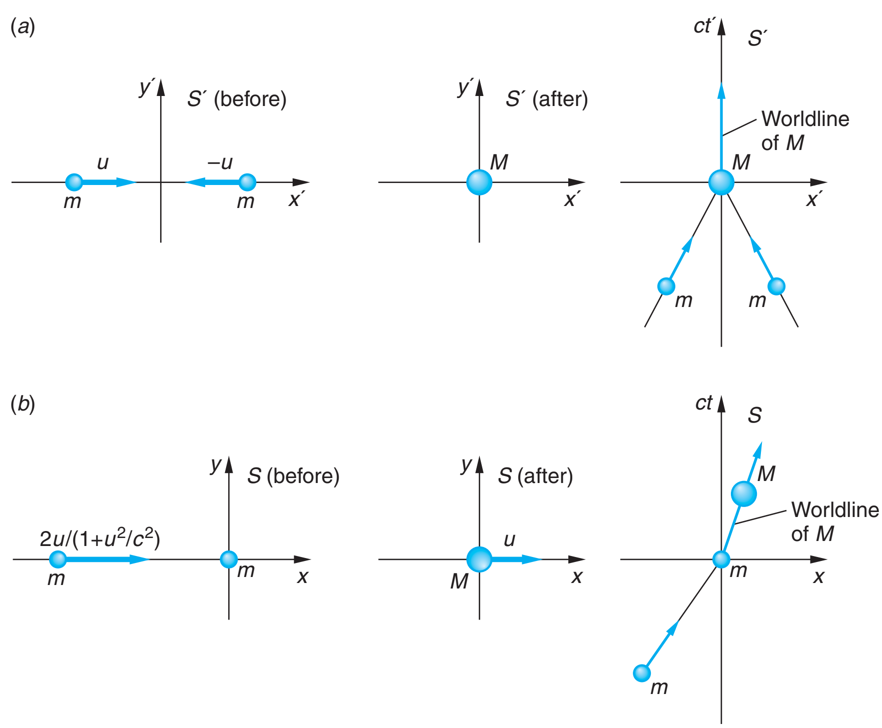


Figure 2-6 Inelastic collision of two particles of equal rest mass m. (*a*) In the zero momentum frame S' the particles have equal and opposite velocities and hence momenta. After the collision, the composite particle of mass M is at rest in S'. The diagram on the far right is the spacetime diagram of the collision from the viewpoint of S'. (*b*) In system S the frame S' is moving to the right at speed u so that the particle on the right is at rest in S, while the left one moves at $2u/(1 + u^2/c^2)$. After collision, the composite particle moves to the right at speed u. Again, the spacetime diagram of the interaction is shown on the far right. All diagrams are drawn with the collision occurring at the origin.

Using the total energy as defined by Equation 2-10, we have in S':

Before collision:

$$E'_{\text{before}} = \frac{mc^2}{\sqrt{1 - u^2/c^2}} + \frac{mc^2}{\sqrt{1 - u^2/c^2}} = \frac{2mc^2}{\sqrt{1 - u^2/c^2}} \tag{2-18}$$

After collision:

$$E'_{\text{after}} = Mc^2 \tag{2-19}$$

Energy will be conserved in S' if $E'_{\text{before}} = E'_{\text{after}}$, i.e., if

$$\frac{2mc^2}{\sqrt{1 - u^2/c^2}} = Mc^2 \tag{2-20}$$

This is ensured by the validity of conservation of momentum, in particular by Equation 2-5, and so energy is conserved in S'. (The validity of Equation 2-20 is important and not trivial. We will consider it in more detail in Example 2-7.) To see if energy as we have defined it is also conserved in S, we transform to S using the inverse transform, Equation 2-17. We then have in S:

Before collision:

$$\begin{aligned} E_{\text{before}} &= \gamma(E'_{\text{before}} + vp'_x) \\ E_{\text{before}} &= \gamma\left(\frac{2mc^2}{\sqrt{1 - u^2/c^2}} + up'_x\right) \\ E_{\text{before}} &= \left(\frac{2mc^2}{\sqrt{1 - u^2/c^2}}\right) \quad \text{since} \quad p'_x = 0 \end{aligned} \tag{2-21}$$

After collision:

$$E_{\text{after}} = \gamma(Mc^2 + up'_x) = \gamma Mc^2 \quad \text{since again} \quad p'_x = 0 \tag{2-22}$$

The energy will be conserved in S and, therefore, the law of conservation of energy will hold in all inertial frames if $E_{\text{before}} = E_{\text{after}}$, i.e., if

$$\gamma\left(\frac{2mc^2}{\sqrt{1 - u^2/c^2}}\right) = \gamma Mc^2 \tag{2-23}$$

which, like Equation 2-20, is ensured by Equation 2-5. Thus, we conclude that the energy as defined by Equation 2-10 is consistent with a relativistically invariant law of conservation of energy, satisfying the first of the conditions set forth at the beginning of this section. While this demonstration was not a general one, since that is beyond the scope of our discussions, you may be assured that our conclusion is quite generally valid.

Question

3. Explain why the result of Example 2-4 does not mean that energy conservation is violated.

EXAMPLE 2-5 Mass of Cosmic Ray Muons In Chapter 1, muons produced as secondary particles by cosmic rays were used to illustrate both the relativistic length contraction and time dilation resulting from their high speed relative to observers on Earth. That speed is about $0.998c$. If the rest energy of a muon is 105.7 MeV, what will observers on Earth measure for the total energy of a cosmic ray–produced muon? What will they measure for its mass?

SOLUTION

The electron volt (eV), the amount of energy acquired by a particle with electric charge equal in magnitude to that on an electron (e) accelerated through a potential difference of 1 volt, is a convenient unit in physics, as you may have learned. It is defined as

$$1.0 \text{ eV} = 1.602 \times 10^{-19} \text{ C} \times 1.0 \text{ V} = 1.602 \times 10^{-19} \text{ J} \qquad \textbf{2-24}$$

Commonly used multiples of the eV are the keV (10^3 eV), the MeV (10^6 eV), the GeV (10^9 eV), and the TeV (10^{12} eV). Many experiments in physics involve the measurement and analysis of the energy and/or momentum of particles and systems of particles, and Equation 2-10 allows us to express the masses of particles in energy units rather than the SI unit of mass, the kilogram. That and the convenient size of the eV facilitate[6] numerous calculations. For example, the mass of an electron is 9.11×10^{-31} kg. Its rest energy is given by

$$E = mc^2 = 9.11 \times 10^{-31} \text{ kg} \cdot c^2 = 8.19 \times 10^{-14} \text{ J}$$

or

$$E = 8.19 \times 10^{-14} \text{ J} \times \frac{1}{1.602 \times 10^{-19} \text{ J/eV}} = 5.11 \times 10^5 \text{ eV}$$

or

$$E = 0.511 \text{ MeV} \quad \textbf{rest energy} \text{ of the electron}$$

The mass of the particle is often expressed with the same number thus:

$$m = \frac{E}{c^2} = 0.511 \text{ MeV}/c^2 \quad \textbf{mass} \text{ of the electron}$$

Now, applying the above to the muons produced by cosmic rays, each has a total energy E given by

$$E = \gamma mc^2 = \frac{1}{\sqrt{1 - (0.998c)^2/c^2}} \times 105.7 \frac{\text{MeV}}{c^2} \times c^2$$
$$E = 1670 \text{ MeV}$$

and a measured mass (see Equation 2-5) of

$$\gamma m = E/c^2 = 1670 \text{ MeV}/c^2$$

The dependence of the measured mass on the speed of the particle has been verified by numerous experiments. Figure 2-7 illustrates a few of those results.

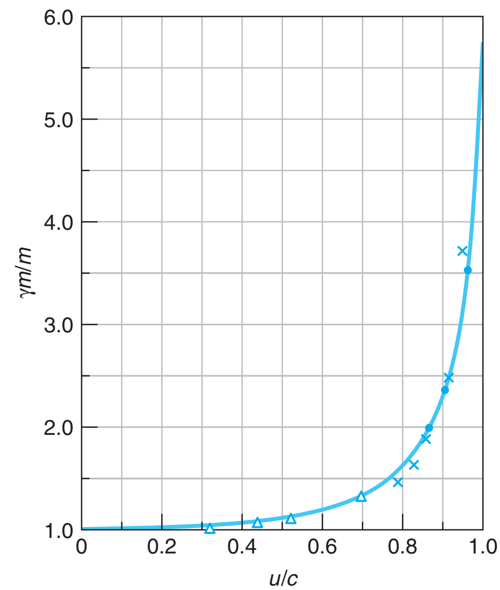


Figure 2-7 A few of the many experimental measurements of the mass of electrons as a function of their speed u/c. The data points are plotted onto Equation 2-5, the solid line. The data points represent the work of W. Kaufmann ($\times$, 1901), A. H. Bucherer (Δ, 1908), and W. Bertozzi (●, 1964). Note that Kaufmann's work preceded the appearance of Einstein's 1905 paper on special relativity. Kaufmann used an incorrect mass for the electron and interpreted his results as support for classical theory. [*Adapted from Figure 3-4 in R. Resnick and D. Halliday, Basic Concepts in Relativity and Early Quantum Theory, 2d ed. (New York: Macmillan, 1992).*]

EXAMPLE 2-6 **Change in the Solar Mass** Compute the rate at which the Sun is losing mass, given that the mean radius R of Earth's orbit is 1.50×10^8 km and the intensity of solar radiation at Earth (called the solar constant) is 1.36×10^3 W/m^2.

SOLUTION

1. The conversion of mass into energy, a consequence of conservation of energy in relativity, is implied by Equation 2-10. With $u = 0$ that equation becomes

$$E = mc^2$$

2. Assuming that the Sun radiates uniformly over a sphere of radius R, the total power radiated by the Sun is given by

$$\begin{aligned} P &= (\text{area of the sphere})(\text{solar constant}) \\ &= (4\pi R^2)(1.36 \times 10^3\ \text{W/m}^2) \\ &= 4\pi(1.50 \times 10^{11}\ \text{m})^2(1.36 \times 10^3\ \text{W/m}^2) \\ &= 3.85 \times 10^{26}\ \text{J/s} \end{aligned}$$

3. Thus, every second the Sun emits 3.85×10^{26} J, which, from Equation 2-10, is the result of converting an amount of mass given by

$$\begin{aligned} m &= E/c^2 \\ &= \frac{3.85 \times 10^{26}\ \text{J}}{(3.00 \times 10^8\ \text{m/s})^2} \\ &= 4.3 \times 10^9\ \text{kg} \end{aligned}$$

Remarks: *Thus, the Sun is losing 4.3×10^9 kg of mass (about 4 million metric tons) every second! If this rate of mass loss remains constant (which it will for the next few billion years) and with a fusion mass-to-energy conversion efficiency of about 1 percent, the Sun's present mass of about 2.0×10^{30} kg will "only" last for about 10^{11} more years!*

EXPLORING

From Mechanics, Another Surprise

One consequence of the fact that Newton's second law, $\mathbf{F} = m\mathbf{a}$, is not relativistically invariant is yet another surprise—the lever paradox. Consider a lever of mass m at rest in S (Figure 2-8). Since the lever is at rest, the net torque τ_{net} due to the forces $\mathbf{F_x}$ and $\mathbf{F_y}$ is zero, i.e. (using magnitudes):

$$\tau_{\text{net}} = \tau_x + \tau_y = -F_x b + F_y a = 0$$

and therefore

$$F_x b = F_y a$$

An observer in system S' moving with $\beta = 0.866$ ($\gamma = 2$) with respect to S sees the lever moving in the $-x'$ direction and measures the torque to be

$$\begin{aligned}\tau'_{\text{net}} &= \tau'_x + \tau'_y = -F'_x b' + F'_y a' = -F_x b + (F_y/2)(a/2)\\ &= -F_x b + F_x b/4 = -(3/4)F_x b \neq 0\end{aligned}$$

where $F'_x = F_x$ and $F'_y = F_y/2$ (see Problem 2-53) and the lever is rotating!

The resolution of the paradox was first given by the German physicist Max von Laue (1879–1960). Recall that the net torque is the rate of change of the angular momentum $\mathbf{L}$. The S' observer measures the work done per unit time by the two forces as

For F'_x: $$-F'_x v = -F_x v$$

For F'_y: zero, since F'_y is perpendicular to the motion

and the change in mass Δm per unit time of the moving lever is

$$\frac{\Delta m}{\Delta t'} = \frac{\Delta E/c^2}{\Delta t'} = \frac{1}{c^2}\frac{\Delta E}{\Delta t'} = -\frac{1}{c^2}F_x v$$

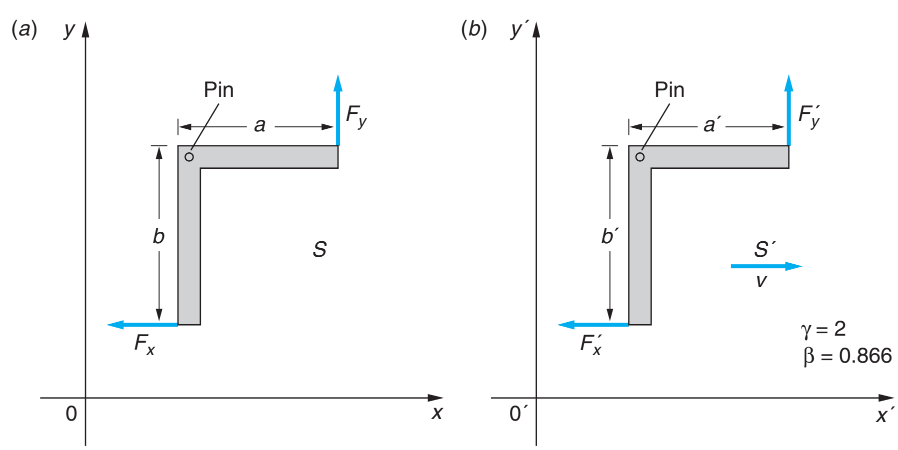


Figure 2-8 (*a*) A lever in the xy plane of system S is free to rotate about the pin P but is held at rest by the two forces F_x and F_y. (*b*) The same lever as seen by an observer in S' that is moving with instantaneous speed v in the $+x$ direction. For the S' observer the lever is moving in the $-x'$ direction.

The S' observer measures a change in the magnitude of angular momentum per unit time given by

$$\tau_{\text{net}} = \frac{\Delta L'}{\Delta t'} = \frac{b\Delta p'}{\Delta t'} = \frac{bv\Delta m}{\Delta t'}$$

Substituting for $\Delta m/\Delta t'$ from above yields

$$\tau'_{\text{net}} = \frac{\Delta L'}{\Delta t'} = bv\frac{-F_x v}{c^2} = -bF_x\frac{v^2}{c^2} = -bF_x\beta^2 = -\frac{3}{4}F_x b$$

As a result of the motion of the lever relative to S', an observer in that system sees the force F'_x doing net work on the lever, thus changing the angular momentum over time, and the paradox vanishes. (The authors thank Costas Efthimiou for bringing this paradox to our attention.)

2-3 Mass/Energy Conversion and Binding Energy

The identification of the term mc^2 as rest energy is not merely a convenience. Whenever additional energy ΔE in any form is stored in an object, the mass of the object is increased by $\Delta E/c^2$. This is of particular importance whenever we want to compare the mass of an object that can be broken into constituent parts with the mass of the parts (for example, an atom containing a nucleus and electrons, or a nucleus containing protons and neutrons). In the case of the atom, the mass changes are usually negligibly small (see Example 2-8). However, the difference between the mass of a *nucleus* and that of its constituent parts (protons and neutrons) is often of great importance.

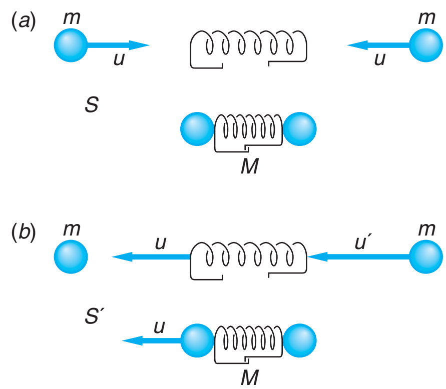


Figure 2-9 Two objects colliding with a massless spring that locks shut. The total rest mass of the system M is greater than that of the parts $2m$ by the amount E_k/c^2, where E_k is the internal energy, which in this case is the original kinetic energy. (*a*) The event as seen in a reference frame S in which the final mass M is at rest. (*b*) The same event as seen in a frame S' moving to the right at speed u relative to S, so that one of the initial masses is at rest.

As an example, consider Figure 2-9*a* in which two particles, each with mass m, are moving toward each other with speeds u. They collide with a spring that compresses and locks shut. (The spring is merely a device for visualizing energy storage.) In the Newtonian mechanics description, the original kinetic energy $E_k = 2\left(\frac{1}{2}mu^2\right)$ is converted into potential energy of the spring U. When the spring is unlocked, the potential energy reappears as kinetic energy of the particles. In relativity theory, the internal energy of the system, $E_k = U$, appears as an increase in the rest mass of the system. That is, the mass of the system M is now greater than $2m$ by E_k/c^2. (We will derive this result in the next example.) This change in mass is too small to be observed for ordinary-size masses and springs, but it is easily observed in transformations that involve nuclei. For example, in the fission of a ^{235}U nucleus, the energy released as kinetic energy of the fission fragments is an appreciable fraction of the rest energy of the original nucleus. (See Example 11-19.) This energy can be calculated by measuring the difference between the mass of the original system and the total mass of the fragments. Einstein was the first to point out this possibility in 1905, even before the discovery of the atomic nucleus, at the end of a very short paper that followed his famous article on relativity.[7] After deriving the theoretical equivalence of energy and mass, he wrote:

The relativistic conversion of mass into energy is the fundamental energy source in the nuclear reactor–based systems that produce electricity in 30 nations and in large naval vessels and nuclear submarines.

> It is not impossible that with bodies whose energy content is variable to a high degree (e.g., with radium salts) the theory may be successfully put to the test.

EXAMPLE 2-7 Change in the Rest Mass of the Two-Particle and Spring System of Figure 2-9 Derive the increase in the rest mass of a system of two particles in a totally inelastic collision. Let m be the mass of each particle so that the total mass of the system is $2m$ when the particles are at rest and far apart, and let M be the rest mass of the system when it has internal energy E_k. The original kinetic energy in the reference frame S (Figure 2-9a) is

$$E_k = 2mc^2(\gamma - 1) \tag{2-25}$$

SOLUTION

In a perfectly inelastic collision, momentum conservation implies that both particles are at rest after collision in this frame, which is the center-of-mass frame. The total kinetic energy is therefore lost. We wish to show that if momentum is to be conserved in any reference frame moving with a constant velocity relative to S, the total mass of the system must increase by Δm, given by

$$\Delta m = \frac{E_k}{c^2} = 2m(\gamma - 1) \tag{2-26}$$

We therefore wish to show that the total mass of the system with internal energy is M, given by

$$M = 2m + \Delta m = 2\gamma m \tag{2-27}$$

To simplify the mathematics, we chose a second reference frame S' moving to the right with speed $v = u$ relative to frame S so that one of the particles is initially at rest, as shown in Figure 2-9b. The initial speed of the other particle in this frame is

$$u' = \frac{u - v}{1 - uv/c^2} = \frac{-2u}{1 + u^2/c^2} \tag{2-28}$$

After collision, the particles move together with speed u toward the left (since they are at rest in S). The initial momentum in S' is

$$p_i' = \frac{mu'}{\sqrt{1 - u'^2/c^2}} \quad \text{to the left}$$

The final momentum is

$$p_f' = \frac{Mu}{\sqrt{1 - u^2/c^2}} \quad \text{to the left}$$

Using Equation 2-28 for u', squaring, dividing by c^2, and adding -1 to both sides gives

$$1 - \frac{u'^2}{c^2} = 1 - \frac{4u^2/c^2}{(1 + u^2/c^2)^2} = \frac{(1 - u^2/c^2)^2}{(1 + u^2/c^2)^2}$$

Then

$$p_i' = \frac{m[2u/(1 + u^2/c^2)]}{(1 - u^2/c^2)/(1 + u^2/c^2)} = \frac{2mu}{(1 - u^2/c^2)}$$

Conservation of momentum in frame S' requires that $p'_f = p'_i$, or

$$\frac{Mu}{\sqrt{1 - u^2/c^2}} = \frac{2mu}{1 - u^2/c^2}$$

Solving for M, we obtain

$$M = \frac{2m}{\sqrt{1 - u^2/c^2}} = 2\gamma m$$

which is Equation 2-27. Thus, the measured value of M would be $2\gamma m$.

If the latch in Figure 2-9*b* were to suddenly come unhooked, the two particles would fly apart with equal momenta, converting the rest mass Δm back into kinetic energy. The derivation is similar to that in Example 2-7.

Mass and Binding Energy

When a system of particles is held together by attractive forces, energy is required to break up the system and separate the particles. The magnitude of this energy E_b is called the *binding energy* of the system. An important result of the special theory of relativity that we will illustrate by example in this section is

The mass of a bound system is less than that of the separated particles by E_b/c^2, where E_b is the binding energy.

In atomic and nuclear physics, masses and energies are typically given in atomic mass units (u) and electron volts (eV) rather than in standard SI units of kilograms and joules. The u is related to the corresponding SI units by

$$1\text{ u} = 1.66054 \times 10^{-27}\text{ kg} = 931.494\text{ MeV}/c^2 \qquad \textbf{2-29}$$

(The eV was defined in terms of the joule in Equation 2-24.) The rest energies of some elementary particles and a few light nuclei are given in Table 2-1, from which you can see by comparing the sums of the masses of the constituent particles with the nuclei listed that the mass of a nucleus is not the same as the sum of the masses of its parts.

The simplest example of nuclear binding energy is that of the deuteron, ^{2}H, which consists of a neutron and a proton bound together. Its rest energy is 1875.613 MeV. The sum of the rest energies of the proton and neutron is 938.272 + 939.565 = 1877.837 MeV. Since this is greater than the rest energy of the deuteron, the deuteron cannot *spontaneously* break up into a neutron and a proton without violating conservation of energy. The binding energy of the deuteron is 1877.837 − 1875.613 = 2.224 MeV. In order to break up a deuteron into a proton and a neutron, at least 2.224 MeV must be added. This can be done by bombarding deuterons with energetic particles or electromagnetic radiation. If a deuteron is formed by combination of a neutron and a proton (fusion; see Chapter 11), the same amount of energy is released.

Table 2-1 Rest energies of some elementary particles and light nuclei

Particle	Symbol	Rest energy (MeV)
Photon	γ	0
Neutrino (antineutrino)	ν ($\bar{\nu}$)	$<2.8 \times 10^{-6}$
Electron (positron)	e or e^- (e^+)	0.5110
Muon	μ^- μ^+	105.7
Pi mesons (pions)	π^- (π^0) π^+	139.6 (135) 139.6
Proton	p	938.272
Neutron	n	939.565
Deuteron	^{2}H or d	1875.613
Helion	^{3}He or h	2808.391
Alpha	^{4}He or α	3727.379

EXAMPLE 2-8 Binding Energy of the Hydrogen Atom The binding energies of atomic electrons to the nuclei of atoms are typically of the order of 10^{-6} times those characteristic of particles in the nuclei; consequently, the mass differences are correspondingly smaller. The binding energy of the hydrogen atom (the energy needed to remove the electron from the atom) is 13.6 eV. How much mass is lost when an electron and a proton form a hydrogen atom?

SOLUTION

The mass of the proton plus that of the electron must be greater than that of the hydrogen atom by

$$\frac{13.6 \text{ eV}}{931.5 \text{ MeV/u}} = 1.46 \times 10^{-8} \text{ u}$$

This mass difference is so small that it is usually neglected.

2-4 Invariant Mass

In Chapter 1 we discovered that as a consequence of Einstein's relativity postulates, the coordinates for space and time are linearly dependent on one another in the Lorentz transformation that connects measurements made in different inertial reference frames. Thus, the time t became a coordinate, in addition to the space coordinates x, y, and z, in the four-dimensional relativistic "world" that we call spacetime. We note in passing that the geometry of spacetime was not the familiar Euclidean geometry of our three-dimensional world, but the four-dimensional Lorentz geometry. The difference became apparent when one compared the computation of the distance r between two points in space with that of the interval between two events in spacetime. The former is, of course, the vector **r**, whose magnitude is given by $r^2 = x^2 + y^2 + z^2$. The vector **r** is unchanged (invariant) under a Galilean transformation in space, and quantities that transform like **r** are also vectors. The latter we called the spacetime interval Δs, and its magnitude, as we have seen, is given by

$$(\Delta s)^2 = (c\Delta t)^2 - [(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2] \qquad \textbf{2-30}$$

The interval Δs is the four-dimensional analog of $\mathbf{r}$ and therefore is called a *four-vector*. Just as x, y, and z are the components of the three-vector $\mathbf{r}$, the components of the four-vector Δs are Δx, Δy, Δz, and $c\Delta t$. We have seen that Δs is also invariant under a Lorentz transformation in spacetime. Correspondingly, any quantity that transforms like Δs—i.e., is invariant under a Lorentz transformation—will also be a four-vector. The physical significance of the invariant interval Δs is quite profound: for timelike intervals $\Delta s/c = \tau$ (the proper time interval), for spacelike intervals $\Delta s = L_p$ (the proper length), and the proper intervals can be found from measurements made in *any* inertial frame.[8]

In the relativistic energy and momentum we have components of another four-vector. In the preceding sections we saw that the momentum and energy, defined by Equations 2-6 and 2-10, respectively, were not only both conserved in relativity, but also together satisfied the Lorentz transformation, Equations 2-16 and 2-17, with the components of the momentum $\mathbf{p}(p_x, p_y, p_z)$ transforming like the space components of $\mathbf{r}(x, y, z)$ and the energy transforming like the time t. The questions then are, What invariant four-vectors are they components of? and, What is its physical significance? The answers to both turn out to be easy to find and yield for us yet another relativistic surprise. By squaring Equations 2-6 and 2-10, you can readily verify that

$$E^2 = (pc)^2 + (mc^2)^2 \qquad \textbf{2-31}$$

This very useful relation we will rearrange slightly to

$$(mc^2)^2 = E^2 - (pc)^2 \qquad \textbf{2-32}$$

Comparing the form of Equation 2-32 with that of Equation 2-30 and knowing that E and $\mathbf{p}$ transform according to the Lorentz transformation, we see that *the magnitude of the invariant energy/momentum four-vector is the rest energy of the mass m*! Thus, observers in all inertial frames will measure the same value for the rest energy of isolated systems and, since c is constant, the same value for the mass. Note that only in the rest frame of the mass m, i.e., the frame where $\mathbf{p} = 0$, are the rest energy and the total energy equal. Even though we have written Equation 2-31 for a single particle, we could as well have written the equations for momentum and energy in terms of the total momentum and total energy of an entire ensemble of noninteracting particles with arbitrary velocities. We would only need to write down Equations 2-6 and 2-10 for each particle and add them together. Thus, the Lorentz transformation for momentum and energy, Equations 2-16 and 2-17, holds for any system of particles and so therefore does the invariance of the rest energy expressed by Equation 2-32.

We may state all of this more formally by saying that the *kinematic* state of the system is described by the four-vector Δs where

$$(\Delta s)^2 = (c\Delta t)^2 - [(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2]$$

and its *dynamic* state is described by the energy/momentum four-vector mc^2, given by

$$(mc^2)^2 = E^2 - (pc)^2$$

The next example illustrates how this works.

EXAMPLE 2-9 Rest Mass of Moving Object A particular object is observed to move through the laboratory at high speed. Its total energy and the components of its momentum are measured by lab workers to be (in SI units) $E = 4.5 \times 10^{17}$ J, $p_x = 3.8 \times 10^8$ kg·m/s, $p_y = 3.0 \times 10^8$ kg·m/s, and $p_z = 3.0 \times 10^8$ kg·m/s. What is the object's rest mass?

SOLUTION A

From Equation 2-32 we can write

$$\begin{aligned}
(mc^2)^2 &= (4.5 \times 10^{17})^2 - [(3.8 \times 10^8 c)^2 + (3.0 \times 10^8 c)^2 + (3.0 \times 10^8 c)^2] \\
&= (4.5 \times 10^{17})^2 - [1.4 \times 10^{17} + 9.0 \times 10^{16} + 9.0 \times 10^{16}]c^2 \\
&= 2.0 \times 10^{35} - 2.9 \times 10^{34} \\
&= 1.74 \times 10^{35}\ \mathrm{J}^2 \\
m &= (1.74 \times 10^{35}\ \mathrm{J}^2)^{1/2}/c^2 = 4.6\ \mathrm{kg}
\end{aligned}$$

SOLUTION B

A slightly different but sometimes more convenient calculation that doesn't involve carrying along large exponentials makes use of Equation 2-32 divided by c^4:

$$m^2 = \left(\frac{E}{c^2}\right)^2 - \left(\frac{p}{c}\right)^2 \qquad \textbf{2-33}$$

Notice that this is simply a unit conversion, expressing each term in (mass)^2 units—e.g., kg^2 when E and p are in SI units:

$$\begin{aligned}
m^2 &= \left(\frac{4.5 \times 10^{17}}{c^2}\right)^2 - \left[\left(\frac{3.8 \times 10^8}{c}\right)^2 + \left(\frac{3.0 \times 10^8}{c}\right)^2 + \left(\frac{3.0 \times 10^8}{c}\right)^2\right] \\
&= (5.0)^2 - [(1.25)^2 + (1.0)^2 + (1.0)^2] \\
&= 25 - 3.56 \\
m &= (21.4)^{1/2} = 4.6\ \mathrm{kg}
\end{aligned}$$

In the example, we determined the rest energy and mass of a rapidly moving object using measurements made in the laboratory without the need to be in the system in which the object was at rest. This ability is of enormous benefit to nuclear, particle, and astrophysicists, whose work regularly involves particles moving at speeds close to that of light. For particles or objects whose rest mass is known, we can use the invariant magnitude of the energy/momentum four-vector to determine the values of other dynamic variables, as illustrated in the next example.

EXAMPLE 2-10 Speed of a Fast Electron The total energy of an electron produced in a particular nuclear reaction is measured to be 2.40 MeV. Find the electron's momentum and speed in the laboratory frame. (The rest mass of an electron is 9.11×10^{-31} kg and its rest energy is 0.511 MeV.)

SOLUTION

The magnitude of the momentum follows immediately from Equation 2-31:

$$\begin{aligned}
pc &= \sqrt{E^2 - (mc^2)^2} = \sqrt{(2.40\ \mathrm{MeV})^2 - (0.511\ \mathrm{MeV})^2} \\
&= 2.34\ \mathrm{MeV} \\
p &= 2.34\ \mathrm{MeV}/c
\end{aligned}$$

where we have again made use of the convenience of the eV as an energy unit. The resulting momentum unit MeV/c can be readily converted to SI units by converting the MeV to joules and dividing by c, i.e.,

$$1 \text{ MeV}/c = \frac{1.602 \times 10^{-13} \text{ J}}{2.998 \times 10^{8} \text{ m/s}} = 5.34 \times 10^{-22} \text{ kg} \cdot \text{m/s}$$

Therefore, the conversion to SI units is easily done, if desired, and yields

$$p = 2.34 \text{ MeV}/c \times \frac{5.34 \times 10^{-22} \text{ kg} \cdot \text{m/s}}{1 \text{ MeV}/c}$$

$$p = 1.25 \times 10^{-21} \text{ kg} \cdot \text{m/s}$$

The speed of the particle is obtained by noting from Equation 2-32 or from Equations 2-6 and 2-10 that

$$\frac{u}{c} = \frac{pc}{E} = \frac{2.34 \text{ MeV}}{2.40 \text{ MeV}} = 0.975 \qquad \textbf{2-34}$$

or

$$u = 0.975c$$

It is extremely important to recognize that the invariant rest energy in Equation 2-32 is that of the *system* and that its value is *not* the sum of the rest energies of the particles of which the system is formed, if the particles move relative to one another. Earlier we used numerical examples of the binding energy of atoms and nuclei that illustrated this fact by showing that the masses of the atoms and nuclei were less than the sum of the masses of their constituents by an amount Δmc^2 that equaled the observed binding energy, but those were systems of interacting particles—i.e., there were forces acting between the constituents. A difference exists, even when the particles do not interact. To see this, let us focus our attention on specifically *what* mass is invariant.

Consider two identical noninteracting particles, each of rest mass $m = 4$ kg moving toward each other along the x axis of S with momentum $p_x = 3c \cdot \text{kg}$, as illustrated in Figure 2-10*a*. The energy of each particle, using Equation 2-33, is

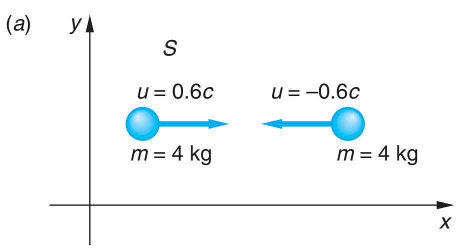
$$E/c^2 = \sqrt{m^2 + (p/c)^2} = \sqrt{(4)^2 + (3)^2} = 5 \text{ kg}$$

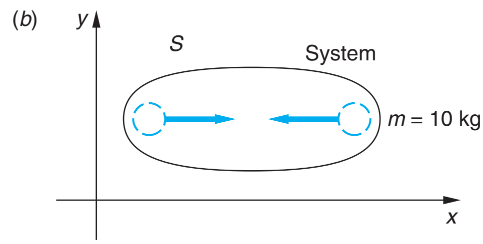
Thus, the total energy of the system is $5c^2 + 5c^2 = 10c^2$ kg, since the energy is a scalar. Similarly, the total momentum of the system is $3c \cdot \text{kg} - 3c \cdot \text{kg} = 0$, since the momentum is a vector and the momenta are equal and opposite. The rest mass of the system is then

$$m = \sqrt{(E/c^2)^2 - (p/c)^2} = \sqrt{(10)^2 - 0^2} = 10 \text{ kg}$$

Hence, the system mass of 10 kg is *greater* than the sum of the masses of the two particles, 8 kg. (This is in contrast to bound systems, such as atoms, where the system mass is *smaller* than the total of the constituents.) This difference is not binding energy, since the particles are noninteracting. Neither does the 2 kg "mass difference" reside equally with the two particles. In fact, it doesn't reside in any particular place but is a property of the entire system. The correct interpretation is that the mass *of the system* is 10 kg.

Although the invariance of the energy/momentum four-vector guarantees that observers in other inertial frames will also measure 10 kg as the mass of the system, let us allow for a skeptic or two and transform to another system S', e.g., the one shown in Figure 2-10*c*, just to be sure.





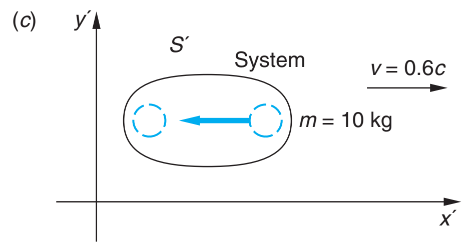


Figure 2-10 (*a*) Two identical particles with rest mass 4 kg approach each other with equal but oppositely directed momenta. The rest mass of the system made up of the two particles is not 4 kg + 4 kg because the system's rest mass includes the mass equivalent of its internal motions. That value, 10 kg (*b*), would be the result of a measurement of the system's mass made by an observer in S, for whom the system is at rest, or by observers in any other inertial frames. (*c*) Transforming to S' moving at $v = 0.6c$ with respect to S, as described in Example 2-11, also yields $m = 10$ kg.

EXAMPLE 2-11 Lorentz Transformation of System Mass For the system illustrated in Figure 2-10, show that an observer in S', which moves relative to S at $\beta = 0.6$, also measures the mass of the system to be 10 kg.

SOLUTION

1. The mass m measured in S' is given by Equation 2-33, which in this case is

$$m = [(E'/c^2)^2 - (p'_x/c)^2]^{1/2}$$

2. E' is given by Equation 2-16:

$$\begin{aligned} E' &= \gamma(E - vp_x) \\ &= \frac{1}{\sqrt{1 - (0.6)^2}}(10c^2 - 0.6c \times 0) \\ &= (1.25)(10c^2) \\ &= 12.5\, c^2 \cdot \text{kg} \end{aligned}$$

3. p'_x is also given by Equation 2-16:

$$\begin{aligned} p'_x &= \gamma(p_x - vE/c^2) \\ &= (1.25)[0 - (0.6c)(10c^2)/c^2] \\ &= -7.5\, c \cdot \text{kg} \end{aligned}$$

4. Substituting E' and p'_x into Equation 2-33 yields

$$\begin{aligned} m &= [(12.5c^2/c^2)^2 - (-7.5c/c)^2]^{1/2} \\ &= [(12.5)^2 - (-7.5)^2]^{1/2} \\ &= 10 \text{ kg} \end{aligned}$$

Remarks: *This result agrees with the value measured in S. The speed of S′ chosen for this calculation, v = 0.6c, is convenient in that one of the particles constituting the system is at rest in S′; however, that has no effect on the generality of the solution.*

Thus, we see that it is the rest energy of any isolated system that is invariant, whether that system is a single atom or the entire universe. And based on our discussions so far, we note that the system's rest energy may be greater than, equal to, or less than the sum of the rest energies of the constituents depending on their relative velocities and the detailed character of any interactions between them.

Questions

4. Suppose two loaded boxcars, each of mass $m = 50$ metric tons, roll toward each other on level track at identical speeds u, collide, and couple together. Discuss the mass of this system before and after the collision. What is the effect of the magnitude of u on your discussion?
5. In 1787 Count Rumford (1753–1814) tried unsuccessfully to measure an increase in the weight of a barrel of water when he increased its temperature from 29°F to 61°F. Explain why, relativistically, you would expect such an increase to occur, and outline an experiment that might, in principle, detect the change. Since Count Rumford preceded Einstein by about 100 years, why might he have been led to such a measurement?

Massless Particles

Equation 2-32 formally allows positive, negative, and zero values for $(mc^2)^2$, just as was the case for the spacetime interval $(\Delta s)^2$. We have been tacitly discussing positive cases so far in this section; a discussion of possible negative cases we will defer until Chapter 12. Here we need to say something about the $mc^2 = 0$ possibility. Note first of all that the idea of zero rest mass has no analog in classical physics since classically $E_k = mu^2/2$ and $\mathbf{p} = m\mathbf{u}$. If $m = 0$, then the momentum and kinetic energy are always zero too, and the "particle" seems to be nothing at all, experiencing no second-law forces, doing no work, and so forth. However, for $mc^2 = 0$, Equation 2-32 states that in relativity

$$E = pc \quad (\text{for } m = 0) \qquad \mathbf{2\text{-}35}$$

and, together with Equation 2-34, that $u = c$, i.e., *a particle whose mass is zero moves at the speed of light*. Similarly, a particle whose speed is measured to be c will have $m = 0$ and satisfy $E = pc$.

We must be careful, however, because Equation 2-32 was obtained from the relativistic definitions of E and $\mathbf{p}$:

$$E = \gamma mc^2 = \frac{mc^2}{\sqrt{1 - u^2/c^2}} \qquad \mathbf{p} = \gamma m\mathbf{u} = \frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}}$$

As $u \to c$, $1/\sqrt{1 - u^2/c^2} \to \infty$; however, since m is also approaching zero, the quantity γm, which is tending toward $0/0$, can (and does) remain defined. Indeed, there is ample experimental evidence for the existence of particles with $mc^2 = 0$.

Current theories suggest the existence of three such particles. Perhaps the most important of these and the one thoroughly verified by experiment is the *photon*, a particle of electromagnetic radiation (i.e., light). Classically, electromagnetic radiation was interpreted via Maxwell's equations as a wave phenomenon, its energy and momentum being distributed continuously throughout the space occupied by the wave.

It was discovered around 1900 that the classical view of light required modification in certain situations, the change being a confinement of the energy and momentum of the radiation into many tiny packets or bundles, which were referred to as photons. Photons move at light speed, of course, and, as we have noted, are required by relativity to have $mc^2 = 0$. Recall that the spacetime interval Δs for light is also zero. Strictly speaking, of course, the second of Einstein's relativity postulates prevents a Lorentz transformation to the rest system of light since light moves at c relative to all inertial frames. Consequently, the term *rest mass* has no operational meaning for light.

EXAMPLE 2-12 Rest Energy of a System of Photons Remember that the rest energy of a system of particles is not the sum of the rest energies of the individual particles if they move relative to one another. This applies to photons too! Suppose two photons, one with energy 5 MeV and the second with energy 2 MeV, approach each other along the x axis. What is the rest energy of this system?

SOLUTION

The momentum of the 5-MeV photon is (from Equation 2-35) $p_x = 5\text{ MeV}/c$ and that of the 2-MeV photon is $p_x = -2\text{ MeV}/c$. Thus, the energy of the system is $E = 5\text{ MeV} + 2\text{ MeV} = 7\text{ MeV}$ and its momentum is $p = 5\text{ MeV}/c - 2\text{ MeV}/c = 3\text{ MeV}/c$. From Equation 2-32 the system's rest energy is

$$mc^2 = \sqrt{(7\text{ MeV})^2 - (3\text{ MeV})^2} = 6.3\text{ MeV}\,!!$$

A second particle whose rest energy is zero is the *gluon*. This massless particle transmits or carries the strong interaction between *quarks*, which are the "building blocks" of all fundamental particles, including protons and neutrons. The existence of gluons is well established experimentally. We will discuss quarks and gluons further in Chapter 12. Finally, there are strong theoretical reasons to expect that gravity is transmitted by a massless particle called the *graviton*, which is related to gravity in much the same way that the photon is related to the electromagnetic field. Gravitons, too, move at speed c. While direct detection of the graviton is beyond our current and foreseeable experimental capabilities, major international cooperative experiments are currently under way to detect gravity waves. (See Section 2-5.)

Until about the beginning of this century a fourth particle, the *neutrino*, was also thought to have zero rest mass. However, substantial experimental evidence collected by the Super-Kamiokande (Japan) and SNO (Canada) imaging neutrino detectors, among others, made it clear that neutrinos are not massless. We discuss neutrino mass and its implications further in Chapters 11 and 12.

Creation and Annihilation of Particles

The relativistic equivalence of mass and energy implies still another remarkable prediction that has no classical counterpart. As long as momentum and energy are conserved in the process,[9] elementary particles with mass can combine with their *antiparticles*, the masses of both being completely converted to energy in a process called *annihilation*. An example is that of an ordinary electron. An electron can orbit briefly with its antiparticle, called a *positron*,[10] but then the two unite, mutually annihilating and producing two or three photons. The two-photon version of this process is shown schematically in Figure 2-11. Positrons are produced naturally by cosmic rays in the upper atmosphere and as the result of the decay of certain radioactive nuclei. P. A. M. Dirac predicted their existence in 1928 while investigating the invariance of the energy/momentum four-vector.

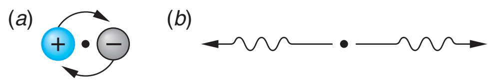
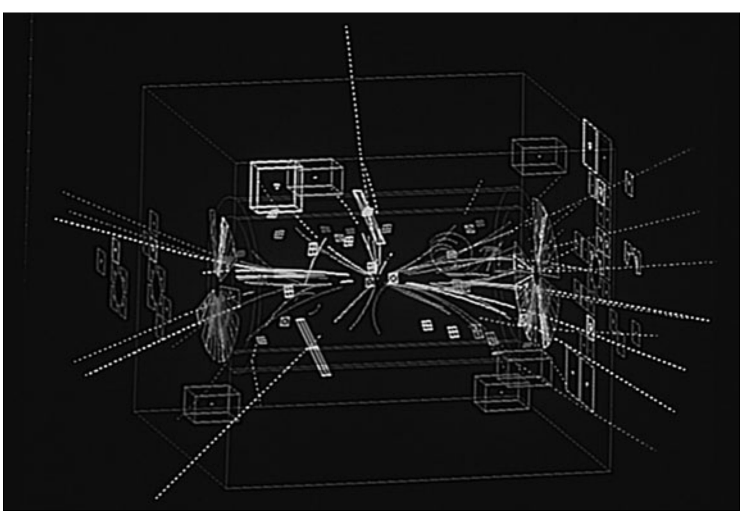


Figure 2-11 (*a*) A positron orbits with an electron about their common center of mass, shown by the dot between them. (*b*) After a short time, typically of the order of 10^{-10} s for the case shown here, the two annihilate, producing two photons. The orbiting electron-positron pair, suggestive of a miniature hydrogen atom, is called *positronium*.

If the speeds of both the electron and the positron $u \ll c$ (not a requirement for the process, but it makes the following calculation clearer), then the total energy of each particle is $E = mc^2 = 0.511$ MeV. Therefore, the total energy of the system in Figure 2-11*a* before annihilation is $2mc^2 = 1.022$ MeV. Also from the diagram, the momenta of the particles are always opposite and equal, and so the total momentum of the system is zero. Conservation of momentum then requires that the total momentum of the two photons produced also be zero, i.e., that they move in opposite directions relative to the original center of mass and have equal momenta. Since $E = pc$ for photons, then they must also have equal energy. Conservation of energy then requires that the energy of each photon be 0.511 MeV. (Photons are usually called *gamma rays* when their energies are a few hundred keV or higher.) Notice from Example 2-12 that the magnitude of the energy/momentum four-vector (the rest energy) is not zero, even though both of the final particles are photons. In this case it equals the rest energy of the initial system. Analysis of the three-photon annihilation, although the calculation is a bit more involved, is similar.

By now you will not be surprised to learn that the reverse process, the creation of mass from energy, can also occur under the proper circumstances. The conversion of mass and energy works both ways. The energy needed to create the new mass can be provided by the kinetic energy of another massive particle or by the "pure" energy of a photon. In either case, in determining what particles might be produced with a given amount of energy, it is important to be sure, as was the case with annihilation, that the appropriate conservation laws are satisfied. As we will discuss in detail in Chapter 12, this restricts the creation process for certain kinds of particles (including electrons, protons, and neutrons) to producing only particle-antiparticle pairs. This means, for example, that the energy in a photon cannot be used to create a single electron but must produce an electron-positron pair.



Decay of a *Z* into an electron-positron pair in the UA1 detectors at CERN. This is the computer image of the first *Z* event recorded (30 April 1983). The newly created pair leave the central detector in opposite directions at nearly the speed of light. [*CERN.*]

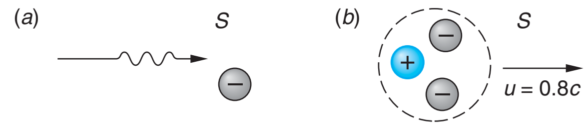


Figure 2-12 (*a*) A photon of energy E and momentum $p = E/c$ encounters an electron at rest. The photon produces an electron-positron pair (*b*), and the group move off together at speed $u = 0.8c$.

To see how relativistic creation of mass goes, let us consider a particular situation, the creation of an electron-positron pair from the energy of a photon. The photon moving through space encounters, or "hits," an electron at rest in frame S, as illustrated in Figure 2-12*a*.[11] Usually the photon simply scatters, but occasionally a pair is created. Encountering the existing electron is important since it is not possible for the photon to produce spontaneously the two rest masses of the pair and also conserve momentum. (See Problem 2-45.) Some other particle must be nearby, not to provide energy to the creation process, but to acquire some of the photon's initial momentum. In this case we have selected an electron for this purpose because it provides a neat example, but almost any particle would do. (See Example 2-13.)

While near the electron, the photon suddenly disappears, and an electron-positron pair appears. The process must occur very fast since the photon, moving at speed c, will travel across a region as large as an atom in about 10^{-19} s. Let's suppose that the details of the interaction that produced the pair are such that the three particles all move off together toward the right in Figure 2-12*b* with the same speed u—i.e., they are all at rest in S', which moves to the right with speed u relative to S.[12] What must the energy E_γ of the photon be for this particular electron-positron pair to be created? To answer this question, we first write the conservation of energy and momentum:

Before pair creation	**After pair creation**
$E_i = E_\gamma + mc^2$	$E_f = E_i = E_\gamma + mc^2$
$p_i = \dfrac{E_\gamma}{c}$	$p_f = p_i = \dfrac{E_\gamma}{c}$

where mc^2 = rest energy of an electron. In the final system after pair creation the total rest energy is $3mc^2$ in this case. We know this because the invariant rest energy equals the sum of the rest energies of the *constituent* particles (the original electron and the pair) in the system where they do not move relative to one another, i.e., in S'. So in S' we have for the system after pair creation

$$(3mc^2)^2 = E^2 - (pc)^2$$

$$9(mc^2)^2 = (E_\gamma + mc^2)^2 - \left(\frac{E_\gamma c}{c}\right)^2$$

$$9(mc^2)^2 = E_\gamma^2 + 2E_\gamma mc^2 + (mc^2)^2 - E_\gamma^2$$

Noting that the E_γ^2 terms cancel and dividing the remaining terms by mc^2, we see that

$$E_\gamma = 4mc^2$$

Thus, the initial photon needs energy equal to 4 electron rest energies in order to create 2 new electron rest masses in this case. Why is the "extra" energy needed? Because the three electrons in the final system share momentum E_γ/c, they must *also* have kinetic energy E_k given by

$$\begin{aligned} E_k &= E - 3mc^2 = (E_\gamma + mc^2) - 3mc^2 \\ &= 4mc^2 + mc^2 - 3mc^2 = 2mc^2 \end{aligned}$$

or the initial photon must provide the $2mc^2$ necessary to create the electron and positron masses and the additional $2mc^2$ of kinetic energy that they and the existing electron share as a result of momentum conservation. The speed u at which the group of particles moves in S can be found from $u/c = pc/E$ (Equation 2-34):

$$u/c = \frac{\left(\frac{E_\gamma}{c} \times c\right)}{(E_\gamma + mc^2)} = \frac{4mc^2}{5mc^2} = 0.8$$

The portion of the incident photon's energy that is needed to provide kinetic energy in the final system is reduced if the mass of the existing particle is larger than that of an electron and, indeed, can be made negligibly small, as illustrated in the following example.

EXAMPLE 2-13 Threshold for Pair Production What is the minimum or threshold energy that a photon must have in order to produce an electron-positron pair?

SOLUTION

The energy E_γ of the initial photon must be

$$E_\gamma = mc^2 + E_{k-} + mc^2 + E_{k+} + E_{kM}$$

where mc^2 = electron rest energy, E_{k-} and E_{k+} are the kinetic energies of the electron and positron, respectively, and E_{kM} = kinetic energy of the existing particle of mass M. Since we are looking for the threshold energy, consider the limiting case where the pair is created at rest in S, i.e., $E_{k-} = E_{k+} = 0$ and correspondingly $p_- = p_+ = 0$. Therefore, momentum conservation requires that

$$p_{\text{initial}} = E_\gamma/c = p_{\text{final}} = \frac{Mu}{\sqrt{1 - u^2/c^2}}$$

where u = speed of recoil of the mass M. Since the masses of single atoms are in the range of 10^3 to 10^5 MeV/c^2 and the value of E_γ at the threshold is clearly less than 2 MeV (i.e., it must be less than the value $E_\gamma = 4mc^2 = 2.044$ MeV, the speed with which M recoils from the creation event is quite small compared with c, even for the smallest M available, a single proton! (See Table 2-1.) Thus, the kinetic energy $E_{kM} \approx \frac{1}{2}mu^2$ becomes negligible, and we conclude that the minimum energy E_γ of the initial photon that can produce an electron-positron pair is $2mc^2$, i.e., that needed just to create the two rest masses.

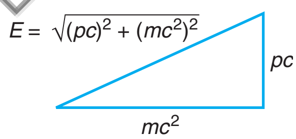


Figure 2-13 Triangle showing the relation between energy, momentum, and rest mass in special relativity. *Caution:* Remember that E and pc are not relativistically invariant. The invariant is mc^2.

Some Useful Equations and Approximations

$$E^2 = (pc)^2 + (mc^2)^2 \tag{2-31}$$

Extremely Relativistic Case The triangle shown in Figure 2-13 is sometimes useful in remembering this result. If the energy of a particle is much greater than its rest energy mc^2, the second term on the right of Equation 2-31 can be neglected, giving the useful approximation

$$E \approx pc \qquad \text{for} \qquad E \gg mc^2 \tag{2-36}$$

This approximation is accurate to about 1 percent or better if E is greater than about $8mc^2$. Equation 2-36 is the exact relation between energy and momentum for particles with zero rest mass.

From Equation 2-36 we see that the momentum of a high-energy particle is simply its total energy divided by c. A convenient unit of momentum is MeV/c. The momentum of a charged particle is usually determined by measuring the radius of curvature of the path of the particle moving in a magnetic field. If the particle has charge q and a velocity $\mathbf{u}$, it experiences a force in a magnetic field $\mathbf{B}$ given by

$$\mathbf{F} = q\mathbf{u} \times \mathbf{B}$$

where $\mathbf{F}$ is perpendicular to the plane formed by $\mathbf{u}$ and $\mathbf{B}$ and hence is always perpendicular to $\mathbf{u}$. Since the magnetic force is always perpendicular to the velocity, it does no work on the particle (the work-energy theorem also holds in relativity), so the energy of the particle is constant. From Equation 2-10 we see that if the energy is constant, γ must be a constant, and therefore the speed u is also constant. So

$$\mathbf{F} = q\mathbf{u} \times \mathbf{B} = \frac{d\mathbf{p}}{dt} = \frac{d(\gamma m\mathbf{u})}{dt} = \gamma m \frac{d\mathbf{u}}{dt}$$

For the case $\mathbf{u} \perp \mathbf{B}$, the particle moves in a circle with centripetal acceleration u^2/R. (If $\mathbf{u}$ is not perpendicular to $\mathbf{B}$, the path is a helix. Since the component of $\mathbf{u}$ parallel to $\mathbf{B}$ is unaffected, we will only consider motion in a plane.) We then have

$$quB = m\gamma\left|\frac{du}{dt}\right| = m\gamma\left(\frac{u^2}{R}\right)$$

or

$$BqR = m\gamma u = p \tag{2-37}$$

This is the same as the nonrelativistic expression except for the factor of γ. Figure 2-14 shows a plot of BqR/mu versus u/c. It is useful to rewrite Equation 2-37 in terms of practical but mixed units; the result is

$$p = 300\, BR\left(\frac{q}{e}\right) \tag{2-38}$$

where p is in MeV/c, B is in tesla, and R is in meters.

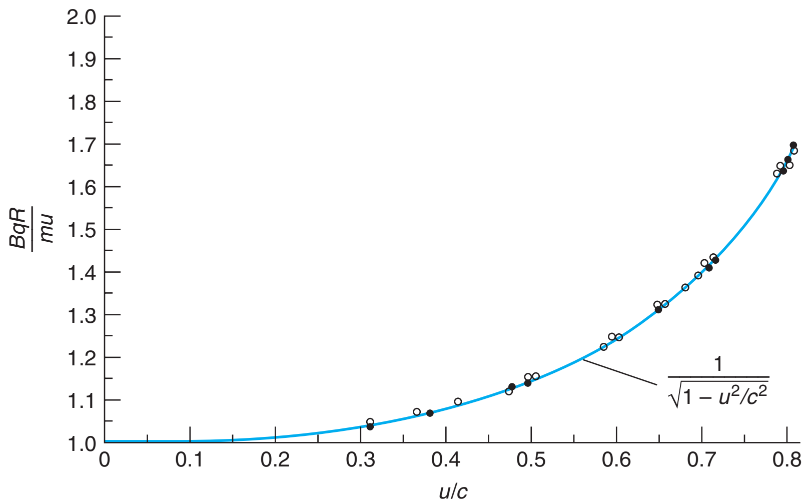


Figure 2-14 BqR/mu versus u/c for particle of charge q and mass m moving in a circular orbit of radius R in a magnetic field B. The agreement of the data with the curve predicted by relativity theory supports the assumption that the force equals the time rate of change of relativistic momentum. [*Adapted from I. Kaplan, Nuclear Physics, 2d ed., Reading, MA: Addison-Wesley Publishing Company, Inc., 1962; by permission.*]

EXAMPLE 2-14 Electron in a Magnetic Field What is the approximate radius of the path of a 30-MeV electron moving in a magnetic field of 0.05 tesla (= 500 gauss)?

SOLUTION

1. The radius of the path is given by rearranging Equation 2-38 and substituting $q = e$:

$$R = \frac{p}{300\, B}$$

2. In this situation the total energy E is much greater than the rest energy mc^2:

$$E = 30 \text{ MeV} \gg mc^2 = 0.511 \text{ MeV}$$

3. Equation 2-36 may then be used to determine p:

$$p \approx E/c = 30 \text{ MeV}/c$$

4. Substituting this approximation for p into Equation 2-38 yields

$$R = \frac{30 \text{ MeV}/c}{(300)(0.05)} = 2 \text{ m}$$

Remarks: *In this case the error made by using the approximation, Equation 2-36, rather than the exact solution, Equation 2-31, is only about 0.01 percent.*

Nonrelativistic Case Nonrelativistic expressions for energy, momentum, and other quantities are often easier to use than the relativistic ones, so it is important to know when these expressions are accurate enough. As $\gamma \to 1$, all the relativistic expressions approach the classical ones. In most situations, the kinetic energy or the total energy is given, so that the most convenient expression for calculating γ is, from Equation 2-10,

$$\gamma = \frac{E}{mc^2} = 1 + \frac{E_k}{mc^2} \qquad \textbf{2-39}$$

When the kinetic energy is much less than the rest energy, γ is approximately 1 and nonrelativistic equations can be used. For example, the classical approximation $E_k \approx (1/2)mu^2 = p^2/2m$ can be used instead of the relativistic $E_k = (\gamma - 1)mc^2$ if E_k is much less than mc^2. We can get an idea of the accuracy of these expressions by expanding γ, using the binomial expansion as was done in Section 2-2, and examining the first term that is *neglected* in the classical approximation. We have

$$\gamma = \left(1 - \frac{u^2}{c^2}\right)^{1/2} \approx 1 + \frac{1}{2}\frac{u^2}{c^2} + \frac{3}{8}\frac{u^4}{c^4} + \cdots$$

and

$$E_k = (\gamma - 1)mc^2 \approx \frac{1}{2}mu^2 + \frac{3}{2}\frac{\left(\frac{1}{2}mu^2\right)^2}{mc^2} + \cdots$$

Then

$$\frac{E_k - \frac{1}{2}mu^2}{E_k} \approx \frac{3}{2}\frac{E_k}{mc^2}$$

For example, if $E_k/mc^2 \approx 1$ percent, the error in using the approximation $E_k \approx (1/2)mu^2$ is about 1.5 percent.

At very low energies, the velocity of a particle can be obtained from its kinetic energy $E_k \approx (1/2)mu^2$ just as in classical mechanics. At very high energies, the velocity of a particle is very near c and the following approximation is sometimes useful (see Problem 2-28):

$$\frac{u}{c} \approx 1 - \frac{1}{2\gamma^2} \qquad \text{for} \quad \gamma \gg 1 \tag{2-40}$$

An exact expression for the velocity of a particle in terms of its energy and momentum was obtained in Example 2-10.

$$\frac{u}{c} = \frac{pc}{E} \tag{2-41}$$

This expression, of course, is not useful if the approximation $E \approx pc$ has already been made.

EXAMPLE 2-15 Different Particles, Same Energy An electron and a proton are each accelerated through 10×10^6 V. Find γ, the momentum, and the speed for each.

SOLUTION

Since each particle has a charge of e, each acquires a kinetic energy of 10 MeV. This is much greater than the 0.511 MeV rest energy of the electron and much less than the 938.3 MeV rest energy of the proton. We will calculate the momentum and speed of each particle exactly and then by means of the nonrelativistic (proton) or the extreme relativistic (electron) approximations.

1. We first consider the electron. From Equation 2-39 we have

$$\gamma = 1 + \frac{E_k}{mc^2} = 1 + \frac{10 \text{ MeV}}{0.511 \text{ MeV}} = 20.57$$