In almost every quantitative problem, the analysis simplifies when you follow the proverbial advice of doing first things first. First approximate and understand the most important effect—the big part—then refine your analysis and understanding. This procedure of successive approximation or “taking out the big part” generates meaningful, memorable, and usable expressions. The following examples introduce the related idea of low-entropy expressions (Section 5.2) and analyze mental multiplication (Section 5.1), exponentiation (Section 5.3), quadratic equations (Section 5.4), and a difficult trigonometric integral (Section 5.5).

### 5.1 Multiplication using one and few

The first illustration is a method of mental multiplication suited to rough, back-of-the-envelope estimates. The particular calculation is the storage capacity of a data CD-ROM. A data CD-ROM has the same format and storage capacity as a music CD, whose capacity can be estimated as the product of three factors:

\[
\text{Storage Capacity} = \frac{1 \text{ hr}}{1 \text{ hr}} \times \frac{3600 \text{ s}}{1 \text{ hr}} \times \frac{4.4 \times 10^4 \text{ samples}}{1 \text{ s}} \times 2 \text{ channels} \times \frac{16 \text{ bits}}{1 \text{ sample}}.
\]

(5.1)
Problem 5.1  Sample rate
Look up the Shannon–Nyquist sampling theorem [22], and explain why the sample rate (the rate at which the sound pressure is measured) is roughly 40 kHz.

Problem 5.2  Bits per sample
Because $2^{16} \sim 10^5$, a 16-bit sample—as chosen for the CD format—requires electronics accurate to roughly 0.001%. Why didn’t the designers of the CD format choose a much larger sample size, say 32 bits (per channel)?

Problem 5.3  Checking units
Check that all the units in the estimate divide out—except for the desired units of bits.

Back-of-the-envelope calculations use rough estimates such as the playing time and neglect important factors such as the bits devoted to error detection and correction. In this and many other estimates, multiplication with 3 decimal places of accuracy would be overkill. An approximate analysis needs an approximate method of calculation.

What is the data capacity to within a factor of 2?

The units (the biggest part!) are bits (Problem 5.3), and the three numerical factors contribute $3600 \times 4.4 \times 10^4 \times 32$. To estimate the product, split it into a big part and a correction.

The big part: The most important factor in a back-of-the-envelope product usually comes from the powers of 10, so evaluate this big part first: 3600 contributes three powers of 10, $4.4 \times 10^4$ contributes four, and 32 contributes one. The eight powers of 10 produce a factor of $10^8$.

The correction: After taking out the big part, the remaining part is a correction factor of $3.6 \times 4.4 \times 3.2$. This product too is simplified by taking out its big part. Round each factor to the closest number among three choices: 1, few, or 10. The invented number few lies midway between 1 and 10: It is the geometric mean of 1 and 10, so $(\text{few})^2 = 10$ and few $\approx 3$. In the product $3.6 \times 4.4 \times 3.2$, each factor rounds to few, so $3.6 \times 4.4 \times 3.2 \approx (\text{few})^3$ or roughly 30.

The units, the powers of 10, and the correction factor combine to give

$$
capacity \sim 10^8 \times 30 \text{ bits} = 3 \times 10^9 \text{ bits}. \tag{5.2}
$$
This estimate is within a factor of 2 of the exact product (Problem 5.4), which is itself close to the actual capacity of $5.6 \times 10^9$ bits.

### Problem 5.4 Underestimate or overestimate?

Does $3 \times 10^9$ overestimate or underestimate $3600 \times 4.4 \times 10^4 \times 32$? Check your reasoning by computing the exact product.

### Problem 5.5 More practice

Use the one-or-few method of multiplication to perform the following calculations mentally; then compare the approximate and actual products.

a. $161 \times 294 \times 280 \times 438$. The actual product is roughly $5.8 \times 10^9$.

b. Earth’s surface area $A = 4\pi R^2$, where the radius is $R \sim 6 \times 10^6$ m. The actual surface area is roughly $5.1 \times 10^{14}$ m$^2$.

## 5.2 Fractional changes and low-entropy expressions

Using the one-or-few method for mental multiplication is fast. For example, $3.15 \times 7.21$ quickly becomes few $\times 10^1 \sim 30$, which is within 50% of the exact product 22.7115. To get a more accurate estimate, round 3.15 to 3 and 7.21 to 7. Their product 21 is in error by only 8%. To reduce the error further, one could split $3.15 \times 7.21$ into a big part and an additive correction. This decomposition produces

$$
(3 + 0.15)(7 + 0.21) = \underbrace{3 \times 7}_{\text{big part}} + \underbrace{0.15 \times 7}_{\text{additive correction}} + \underbrace{3 \times 0.21}_{\text{additive correction}} + \underbrace{0.15 \times 0.21}_{\text{additive correction}}.
$$

The approach is sound, but the literal application of taking out the big part produces a messy correction that is hard to remember and understand. Slightly modified, however, taking out the big part provides a clean and intuitive correction. As gravy, developing the improved correction introduces two important street-fighting ideas: fractional changes (Section 5.2.1) and low-entropy expressions (Section 5.2.2). The improved correction will then, as a first of many uses, help us estimate the energy saved by highway speed limits (Section 5.2.3).

### 5.2.1 Fractional changes

The hygienic alternative to an additive correction is to split the product into a big part and a multiplicative correction:
Taking out the big part

\[ 3.15 \times 7.21 = 3 \times 7 \times (1 + 0.05) \times (1 + 0.03). \]  \hspace{1cm} (5.4)

**Can you find a picture for the correction factor?**

The correction factor is the area of a rectangle with width \(1 + 0.05\) and height \(1 + 0.03\). The rectangle contains one subrectangle for each term in the expansion of \((1 + 0.05) \times (1 + 0.03)\). Their combined area of roughly \(1 + 0.05 + 0.03\) represents an 8% fractional increase over the big part. The big part is 21, and 8% of it is 1.68, so \(3.15 \times 7.21 = 22.68\), which is within 0.14% of the exact product.

**Problem 5.6 Picture for the fractional error**

What is the pictorial explanation for the fractional error of roughly 0.15%?

**Problem 5.7 Try it yourself**

Estimate \(245 \times 42\) by rounding each factor to a nearby multiple of 10, and compare this big part with the exact product. Then draw a rectangle for the correction factor, estimate its area, and correct the big part.

### 5.2.2 Low-entropy expressions

The correction to \(3.15 \times 7.21\) was complicated as an absolute or additive change but simple as a fractional change. This contrast is general. Using the additive correction, a two-factor product becomes

\[
(x + \Delta x)(y + \Delta y) = xy + x\Delta y + y\Delta x + \Delta x\Delta y. \tag{5.5}
\]

**Problem 5.8 Rectangle picture**

Draw a rectangle representing the expansion

\[
(x + \Delta x)(y + \Delta y) = xy + x\Delta y + y\Delta x + \Delta x\Delta y. \tag{5.6}
\]

When the absolute changes \(\Delta x\) and \(\Delta y\) are small \((x \ll \Delta x\) and \(y \ll \Delta y\)), the correction simplifies to \(x\Delta y + y\Delta x\), but even so it is hard to remember because it has many plausible but incorrect alternatives. For example, it could plausibly contain terms such as \(\Delta x\Delta y\), \(x\Delta x\), or \(y\Delta y\). The extent
of the plausible alternatives measures the gap between our intuition and reality; the larger the gap, the harder the correct result must work to fill it, and the harder we must work to remember the correct result.

Such gaps are the subject of statistical mechanics and information theory [20, 21], which define the gap as the logarithm of the number of plausible alternatives and call the logarithmic quantity the entropy. The logarithm does not alter the essential point that expressions differ in the number of plausible alternatives and that high-entropy expressions [28]—ones with many plausible alternatives—are hard to remember and understand.

In contrast, a low-entropy expression allows few plausible alternatives, and elicits, “Yes! How could it be otherwise?!” Much mathematical and scientific progress consists of finding ways of thinking that turn high-entropy expressions into easy-to-understand, low-entropy expressions.

What is a low-entropy expression for the correction to the product $xy$?

A multiplicative correction, being dimensionless, automatically has lower entropy than the additive correction: The set of plausible dimensionless expressions is much smaller than the full set of plausible expressions.

The multiplicative correction is $(x + \Delta x)(y + \Delta y)/xy$. As written, this ratio contains gratuitous entropy. It constructs two dimensioned sums $x + \Delta x$ and $y + \Delta y$, multiplies them, and finally divides the product by $xy$. Although the result is dimensionless, it becomes so only in the last step. A cleaner method is to group related factors by making dimensionless quantities right away:

$$\frac{(x + \Delta x)(y + \Delta y)}{xy} = \frac{x + \Delta x}{x} \frac{y + \Delta y}{y} = \left(1 + \frac{\Delta x}{x}\right) \left(1 + \frac{\Delta y}{y}\right). \quad (5.7)$$

The right side is built only from the fundamental dimensionless quantity 1 and from meaningful dimensionless ratios: $(\Delta x)/x$ is the fractional change in $x$, and $(\Delta y)/y$ is the fractional change in $y$.

The gratuitous entropy came from mixing $x + \Delta x$, $y + \Delta y$, $x$, and $y$ willy nilly, and it was removed by regrouping or unmixing. Unmixing is difficult with physical systems. Try, for example, to remove a drop of food coloring mixed into a glass of water. The problem is that a glass of water contains roughly $10^{25}$ molecules. Fortunately, most mathematical expressions have fewer constituents. We can often regroup and unmix the mingled pieces and thereby reduce the entropy of the expression.
Problem 5.9  Rectangle for the correction factor

Draw a rectangle representing the low-entropy correction factor

\[ \left(1 + \frac{\Delta x}{x}\right)\left(1 + \frac{\Delta y}{y}\right). \] (5.8)

A low-entropy correction factor produces a low-entropy fractional change:

\[ \frac{\Delta(xy)}{xy} = \left(1 + \frac{\Delta x}{x}\right)\left(1 + \frac{\Delta y}{y}\right) - 1 = \frac{\Delta x}{x} + \frac{\Delta y}{y} + \frac{\Delta x}{x} \frac{\Delta y}{y}, \] (5.9)

where \( \Delta(xy)/xy \) is the fractional change from \( xy \) to \((x + \Delta x)(y + \Delta y)\). The rightmost term is the product of two small fractions, so it is small compared to the preceding two terms. Without this small, quadratic term,

\[ \frac{\Delta(xy)}{xy} \approx \frac{\Delta x}{x} + \frac{\Delta y}{y}. \] (5.10)

Small fractional changes simply add!

This fractional-change rule is far simpler than the corresponding approximate rule that the absolute change is \( x\Delta y + y\Delta x \). Simplicity indicates low entropy; indeed, the only plausible alternative to the proposed rule is the possibility that fractional changes multiply. And this conjecture is not likely: When \( \Delta y = 0 \), it predicts that \( \Delta(xy) = 0 \) no matter the value of \( \Delta x \) (this prediction is explored also in Problem 5.12).

Problem 5.10  Thermal expansion

If, due to thermal expansion, a metal sheet expands in each dimension by 4%, what happens to its area?

Problem 5.11  Price rise with a discount

Imagine that inflation, or copyright law, increases the price of a book by 10% compared to last year. Fortunately, as a frequent book buyer, you start getting a store discount of 15%. What is the net price change that you see?

5.2.3  Squaring

In analyzing the engineered and natural worlds, a common operation is squaring—a special case of multiplication. Squared lengths are areas, and squared speeds are proportional to the drag on most objects (Section 2.4):

\[ F_d \sim \rho v^2 A, \] (5.11)
where $v$ is the speed of the object, $A$ is its cross-sectional area, and $\rho$ is the density of the fluid. As a consequence, driving at highway speeds for a distance $d$ consumes an energy $E = F_d d \sim \rho A v^2 d$. Energy consumption can therefore be reduced by driving more slowly. This possibility became important to Western countries in the 1970s when oil prices rose rapidly (see [7] for an analysis). As a result, the United States instituted a highway speed limit of 55 mph (90 kph).

By what fraction does gasoline consumption fall due to driving 55 mph instead of 65 mph?

A lower speed limit reduces gasoline consumption by reducing the drag force $\rho A v^2$ and by reducing the driving distance $d$: People measure and regulate their commuting more by time than by distance. But finding a new home or job is a slow process. Therefore, analyze first things first—assume for this initial analysis that the driving distance $d$ stays fixed (then try Problem 5.14).

With that assumption, $E$ is proportional to $v^2$, and

$$\frac{\Delta E}{E} = 2 \times \frac{\Delta v}{v}.$$  \hspace{1cm} (5.12)

Going from 65 mph to 55 mph is roughly a 15\% drop in $v$, so the energy consumption drops by roughly 30\%. Highway driving uses a significant fraction of the oil consumed by motor vehicles, which in the United States consume a significant fraction of all oil consumed. Thus the 30\% drop substantially reduced total US oil consumption.

**Problem 5.12**  A tempting error

If $A$ and $x$ are related by $A = x^2$, a tempting conjecture is that

$$\frac{\Delta A}{A} \approx \left( \frac{\Delta x}{x} \right)^2.$$ \hspace{1cm} (5.13)

Disprove this conjecture using easy cases (Chapter 2).

**Problem 5.13**  Numerical estimates

Use fractional changes to estimate $6.3^3$. How accurate is the estimate?

**Problem 5.14**  Time limit on commuting

Assume that driving time, rather than distance, stays fixed as highway driving speeds fall by 15\%. What is the resulting fractional change in the gasoline consumed by highway driving?
Problem 5.15  Wind power
The power generated by an ideal wind turbine is proportional to $v^3$ (why?). If wind speeds increase by a mere 10%, what is the effect on the generated power? The quest for fast winds is one reason that wind turbines are placed on cliffs or hilltops or at sea.

5.3 Fractional changes with general exponents

The fractional-change approximations for changes in $x^2$ (Section 5.2.3) and in $x^3$ (Problem 5.13) are special cases of the approximation for $x^n$

$$\frac{\Delta (x^n)}{x^n} \approx n \times \frac{\Delta x}{x}. \quad (5.14)$$

This rule offers a method for mental division (Section 5.3.1), for estimating square roots (Section 5.3.2), and for judging a common explanation for the seasons (Section 5.3.3). The rule requires only that the fractional change be small and that the exponent $n$ not be too large (Section 5.3.4).

5.3.1 Rapid mental division

The special case $n = -1$ provides the method for rapid mental division. As an example, let’s estimate $1/13$. Rewrite it as $(x + \Delta x)^{-1}$ with $x = 10$ and $\Delta x = 3$. The big part is $x^{-1} = 0.1$. Because $(\Delta x)/x = 30\%$, the fractional correction to $x^{-1}$ is roughly $-30\%$. The result is 0.07.

$$\frac{1}{13} \approx \frac{1}{10} - 30\% = 0.07, \quad (5.15)$$

where the “$-30\%$” notation, meaning “decrease the previous object by 30\%,” is a useful shorthand for a factor of $1 - 0.3$.

How accurate is the estimate, and what is the source of the error?

The estimate is in error by only 9%. The error arises because the linear approximation

$$\frac{\Delta (x^{-1})}{x^{-1}} \approx -1 \times \frac{\Delta x}{x} \quad (5.16)$$

does not include the square (or higher powers) of the fractional change $(\Delta x)/x$ (Problem 5.17 asks you to find the squared term).
How can the error in the linear approximation be reduced?

To reduce the error, reduce the fractional change. Because the fractional change is determined by the big part, let’s increase the accuracy of the big part. Accordingly, multiply $1/13$ by $8/8$, a convenient form of $1$, to construct $8/104$. Its big part $0.08$ approximates $1/13$ already to within $4\%$. To improve it, write $1/104$ as $(x + \Delta x)^{-1}$ with $x = 100$ and $\Delta x = 4$. The fractional change $(\Delta x)/x$ is now $0.04$ (rather than $0.3$); and the fractional correction to $1/x$ and $8/x$ is a mere $-4\%$. The corrected estimate is $0.0768$:

$$\frac{1}{13} \approx 0.08 - 4\% = 0.08 - 0.0032 = 0.0768. \quad (5.17)$$

This estimate can be done mentally in seconds and is accurate to $0.13\%$!

**Problem 5.16** Next approximation

Multiply $1/13$ by a convenient form of $1$ to make a denominator near $1000$; then estimate $1/13$. How accurate is the resulting approximation?

**Problem 5.17** Quadratic approximation

Find $A$, the coefficient of the quadratic term in the improved fractional-change approximation

$$\frac{\Delta (x^{-1})}{x^{-1}} \approx -1 \times \frac{\Delta x}{x} + A \times \left( \frac{\Delta x}{x} \right)^2. \quad (5.18)$$

Use the resulting approximation to improve the estimates for $1/13$.

**Problem 5.18** Fuel efficiency

Fuel efficiency is inversely proportional to energy consumption. If a $55$ mph speed limit decreases energy consumption by $30\%$, what is the new fuel efficiency of a car that formerly got $30$ miles per US gallon ($12.8$ kilometers per liter)?

### 5.3.2 Square roots

The fractional exponent $n = 1/2$ provides the method for estimating square roots. As an example, let’s estimate $\sqrt{10}$. Rewrite it as $(x + \Delta x)^{1/2}$ with $x = 9$ and $\Delta x = 1$. The big part $x^{1/2}$ is $3$. Because $(\Delta x)/x = 1/9$ and $n = 1/2$, the fractional correction is $1/18$. The corrected estimate is

$$\sqrt{10} \approx 3 \times \left( 1 + \frac{1}{18} \right) \approx 3.1667. \quad (5.19)$$

The exact value is $3.1622\ldots$, so the estimate is accurate to $0.14\%$. 

Problem 5.19 Overestimate or underestimate?
Does the linear fractional-change approximation overestimate all square roots (as it overestimated $\sqrt{10}$)? If yes, explain why; if no, give a counterexample.

Problem 5.20 Cosine approximation
Use the small-angle approximation $\sin \theta \approx \theta$ to show that $\cos \theta \approx 1 - \theta^2/2$.

Problem 5.21 Reducing the fractional change
To reduce the fractional change when estimating $\sqrt{10}$, rewrite it as $\sqrt{360}/6$ and then estimate $\sqrt{360}$. How accurate is the resulting estimate for $\sqrt{10}$?

Problem 5.22 Another method to reduce the fractional change
Because $\sqrt{2}$ is fractionally distant from the nearest integer square roots $\sqrt{1}$ and $\sqrt{4}$, fractional changes do not give a direct and accurate estimate of $\sqrt{2}$. A similar problem occurred in estimating $\ln 2$ (Section 4.3); there, rewriting $2$ as $(4/3)/(2/3)$ improved the accuracy. Does that rewriting help estimate $\sqrt{2}$?

Problem 5.23 Cube root
Estimate $2^{1/3}$ to within 10%.

5.3.3 A reason for the seasons?

Summers are warmer than winters, it is often alleged, because the earth is closer to the sun in the summer than in the winter. This common explanation is bogus for two reasons. First, summers in the southern hemisphere happen alongside winters in the northern hemisphere, despite almost no difference in the respective distances to the sun. Second, as we will now estimate, the varying earth–sun distance produces too small a temperature difference. The causal chain—that the distance determines the intensity of solar radiation and that the intensity determines the surface temperature—is most easily analyzed using fractional changes.

Intensity of solar radiation: The intensity is the solar power divided by the area over which it spreads. The solar power hardly changes over a year (the sun has existed for several billion years); however, at a distance $r$ from the sun, the energy has spread over a giant sphere with surface area $\sim r^2$. The intensity $I$ therefore varies according to $I \propto r^{-2}$. The fractional changes in radius and intensity are related by

$$\frac{\Delta I}{I} \approx -2 \times \frac{\Delta r}{r}.$$

(5.20)
Surface temperature: The incoming solar energy cannot accumulate and returns to space as blackbody radiation. Its outgoing intensity depends on the earth’s surface temperature $T$ according to the Stefan–Boltzmann law $I = \sigma T^4$ (Problem 1.12), where $\sigma$ is the Stefan–Boltzmann constant. Therefore $T \propto I^{1/4}$. Using fractional changes,

$$\frac{\Delta T}{T} \approx \frac{1}{4} \times \frac{\Delta I}{I}.$$  \hfill (5.21)

This relation connects intensity and temperature. The temperature and distance are connected by $(\Delta I)/I = -2 \times (\Delta r)/r$. When joined, the two relations connect distance and temperature as follows:

\[
\begin{align*}
\frac{\Delta r}{r} & \quad \rightarrow \quad -2 \quad \rightarrow \quad \frac{\Delta I}{I} \approx -2 \times \frac{\Delta r}{r} \quad \rightarrow \quad \frac{\Delta T}{T} \approx -\frac{1}{2} \times \frac{\Delta r}{r}
\end{align*}
\]

The next step in the computation is to estimate the input $(\Delta r)/r$—namely, the fractional change in the earth–sun distance. The earth orbits the sun in an ellipse; its orbital distance is

$$r = \frac{l}{1 + \epsilon \cos \theta},$$  \hfill (5.22)

where $\epsilon$ is the eccentricity of the orbit, $\theta$ is the polar angle, and $l$ is the semilatus rectum. Thus $r$ varies from $r_{\text{min}} = l/(1 + \epsilon)$ (when $\theta = 0^\circ$) to $r_{\text{max}} = l/(1 - \epsilon)$ (when $\theta = 180^\circ$). The increase from $r_{\text{min}}$ to $l$ contributes a fractional change of roughly $\epsilon$. The increase from $l$ to $r_{\text{max}}$ contributes another fractional change of roughly $\epsilon$. Thus, $r$ varies by roughly $2\epsilon$. For the earth’s orbit, $\epsilon = 0.016$, so the earth–sun distance varies by 0.032 or 3.2% (making the intensity vary by 6.4%).

**Problem 5.24  Where is the sun?**

The preceding diagram of the earth’s orbit placed the sun away from the center of the ellipse. The diagram to the right shows the sun at an alternative and perhaps more natural location: at the center of the ellipse. What physical laws, if any, prevent the sun from sitting at the center of the ellipse?

**Problem 5.25  Check the fractional change**

Look up the minimum and maximum earth–sun distances and check that the distance does vary by 3.2% from minimum to maximum.
A 3.2\% increase in distance causes a slight drop in temperature:

\[
\frac{\Delta T}{T} \approx -\frac{1}{2} \times \frac{\Delta r}{r} = -1.6\%. \tag{5.23}
\]

However, man does not live by fractional changes alone and experiences the absolute temperature change \(\Delta T\).

\[
\Delta T = -1.6\% \times T. \tag{5.24}
\]

In winter \(T \approx 0^\circ C\), so is \(\Delta T \approx 0^\circ C\)?

If our calculation predicts that \(\Delta T \approx 0^\circ C\), it must be wrong. An even less plausible conclusion results from measuring \(T\) in Fahrenheit degrees, which makes \(T\) often negative in parts of the northern hemisphere. Yet \(\Delta T\) cannot flip its sign just because \(T\) is measured in Fahrenheit degrees!

Fortunately, the temperature scale is constrained by the Stefan–Boltzmann law. For blackbody flux to be proportional to \(T^4\), temperature must be measured relative to a state with zero thermal energy: absolute zero. Neither the Celsius nor the Fahrenheit scale satisfies this requirement.

In contrast, the Kelvin scale does measure temperature relative to absolute zero. On the Kelvin scale, the average surface temperature is \(T \approx 300\, K\); thus, a 1.6\% change in \(T\) makes \(\Delta T \approx 5\, K\). A 5\, K change is also a 5\, ^\circ C change—Kelvin and Celsius degrees are the same size, although the scales have different zero points. (See also Problem 5.26.) A typical temperature change between summer and winter in temperate latitudes is 20\, ^\circ C—much larger than the predicted 5\, ^\circ C change, even after allowing for errors in the estimate. A varying earth–sun distance is a dubious explanation of the reason for the seasons.

**Problem 5.26 Converting to Fahrenheit**

The conversion between Fahrenheit and Celsius temperatures is

\[
F = 1.8C + 32, \tag{5.25}
\]

so a change of 5\, ^\circ C should be a change of 41\, ^\circ F—sufficiently large to explain the seasons! What is wrong with this reasoning?

**Problem 5.27 Alternative explanation**

If a varying distance to the sun cannot explain the seasons, what can? Your proposal should, in passing, explain why the northern and southern hemispheres have summer 6 months apart.
5.3.4 Limits of validity

The linear fractional-change approximation

$$\frac{\Delta (x^n)}{x^n} \approx n \times \frac{\Delta x}{x} \quad (5.26)$$

has been useful. But when is it valid? To investigate without drowning in notation, write $z$ for $\Delta x$; then choose $x = 1$ to make $z$ the absolute and the fractional change. The right side becomes $nz$, and the linear fractional-change approximation is equivalent to

$$(1 + z)^n \approx 1 + nz. \quad (5.27)$$

The approximation becomes inaccurate when $z$ is too large: for example, when evaluating $\sqrt{1+z}$ with $z = 1$ (Problem 5.22). Is the exponent $n$ also restricted? The preceding examples illustrated only moderate-sized exponents: $n = 2$ for energy consumption (Section 5.2.3), $-2$ for fuel efficiency (Problem 5.18), $-1$ for reciprocals (Section 5.3.1), $1/2$ for square roots (Section 5.3.2), and $-2$ and $1/4$ for the seasons (Section 5.3.3). We need further data.

What happens in the extreme case of large exponents?

With a large exponent such as $n = 100$ and, say, $z = 0.001$, the approximation predicts that $1.001^{100} \approx 1.1$—close to the true value of 1.105… However, choosing the same $n$ alongside $z = 0.1$ (larger than 0.001 but still small) produces the terrible prediction

$$\underbrace{1.1^{100}}_{\text{nz}} = 1 + 100 \times 0.1 = 11; \quad (5.28)$$

$1.1^{100}$ is roughly 14,000, more than 1000 times larger than the prediction. Both predictions used large $n$ and small $z$, yet only one prediction was accurate; thus, the problem cannot lie in $n$ or $z$ alone. Perhaps the culprit is the dimensionless product $nz$. To test that idea, hold $nz$ constant while trying large values of $n$. For $nz$, a sensible constant is 1—the simplest dimensionless number. Here are several examples.

$$1.1^{10} \approx 2.59374,$$

$$1.01^{100} \approx 2.70481, \quad (5.29)$$

$$1.001^{1000} \approx 2.71692.$$
In each example, the approximation incorrectly predicts that \((1+z)^n = 2\).

What is the cause of the error?

To find the cause, continue the sequence beyond 1.001\(^{1000}\) and hope that a pattern will emerge: The values seem to approach \(e = 2.718281828\ldots\), the base of the natural logarithms. Therefore, take the logarithm of the whole approximation.

\[
\ln(1+z)^n = n \ln(1+z). \tag{5.30}
\]

Pictorial reasoning showed that \(\ln(1+z) \approx z\) when \(z \ll 1\) (Section 4.3). Thus, \(n \ln(1+z) \approx nz\), making \((1+z)^n \approx e^{nz}\). This improved approximation explains why the approximation \((1+z)^n \approx 1 + nz\) failed with large \(nz\): Only when \(nz \ll 1\) is \(e^{nz}\) approximately 1 + nz. Therefore, when \(z \ll 1\) the two simplest approximation are

\[
(1+z)^n \approx \begin{cases} 
1 + nz & (z \ll 1 \text{ and } nz \ll 1), \\
e^{nz} & (z \ll 1 \text{ and } nz \text{ unrestricted}).
\end{cases} \tag{5.31}
\]

The diagram shows, across the whole \(n-z\) plane, the simplest approximation in each region. The axes are logarithmic and \(n\) and \(z\) are assumed positive: The right half plane shows \(z \gg 1\), and the upper half plane shows \(n \gg 1\). On the lower right, the boundary curve is \(n \ln z = 1\). Explaining the boundaries and extending the approximations is an instructive exercise (Problem 5.28).

**Problem 5.28 Explaining the approximation plane**

In the right half plane, explain the \(n/z = 1\) and \(n \ln z = 1\) boundaries. For the whole plane, relax the assumption of positive \(n\) and \(z\) as far as possible.

**Problem 5.29 Binomial-theorem derivation**

Try the following alternative derivation of \((1+z)^n \approx e^{nz}\) (where \(n \gg 1\)). Expand \((1+z)^n\) using the binomial theorem, simplify the products in the binomial coefficients by approximating \(n-k\) as \(n\), and compare the resulting expansion to the Taylor series for \(e^{nz}\).
5.4 Successive approximation: How deep is the well?

The next illustration of taking out the big part emphasizes successive approximation and is disguised as a physics problem.

You drop a stone down a well of unknown depth \( h \) and hear the splash 4 s later. Neglecting air resistance, find \( h \) to within 5%. Use \( c_s = 340 \text{ m s}^{-1} \) as the speed of sound and \( g = 10 \text{ m s}^{-2} \) as the strength of gravity.

Approximate and exact solutions give almost the same well depth, but offer significantly different understandings.

5.4.1 Exact depth

The depth is determined by the constraint that the 4 s wait splits into two times: the rock falling freely down the well and the sound traveling up the well. The free-fall time is \( \sqrt{\frac{2h}{g}} \) (Problem 1.3), so the total time is

\[
T = \sqrt{\frac{2h}{g}} + \frac{h}{c_s}.
\] (5.32)

To solve for \( h \) exactly, either isolate the square root on one side and square both sides to get a quadratic equation in \( h \) (Problem 5.30); or, for a less error-prone method, rewrite the constraint as a quadratic equation in a new variable \( z = \sqrt{h} \).

**Problem 5.30 Other quadratic**

Solve for \( h \) by isolating the square root on one side and squaring both sides. What are the advantages and disadvantages of this method in comparison with the method of rewriting the constraint as a quadratic in \( z = \sqrt{h} \)?

As a quadratic equation in \( z = \sqrt{h} \), the constraint is

\[
\frac{1}{c_s} z^2 + \sqrt{\frac{2}{g}} z - T = 0.
\] (5.33)

Using the quadratic formula and choosing the positive root yields

\[
z = \frac{-\sqrt{2/g}}{2/c_s} + \frac{\sqrt{2/g + 4T/c_s}}{2/c_s}.
\] (5.34)

Because \( z^2 = h \),
\[ h = \left( \frac{-\sqrt{2/g} + \sqrt{2/g + 4T/c_s}}{2/c_s} \right)^2. \] (5.35)

Substituting \( g = 10 \text{ m s}^{-2} \) and \( c_s = 340 \text{ m s}^{-1} \) gives \( h \approx 71.56 \text{ m} \).

Even if the depth is correct, the exact formula for it is a mess. Such high-entropy horrors arise frequently from the quadratic formula; its use often signals the triumph of symbol manipulation over thought. Exact answers, we will find, may be less useful than approximate answers.

### 5.4.2 Approximate depth

To find a low-entropy, approximate depth, identify the big part—the most important effect. Here, most of the total time is the rock’s free fall: The rock’s maximum speed, even if it fell for the entire 4 s, is only \( gT = 40 \text{ m s}^{-1} \), which is far below \( c_s \). Therefore, the most important effect should arise in the extreme case of infinite sound speed.

**If \( c_s = \infty \), how deep is the well?**

In this zeroth approximation, the free-fall time \( t_0 \) is the full time \( T = 4 \text{ s} \), so the well depth \( h_0 \) becomes

\[ h_0 = \frac{1}{2} gt_0^2 = 80 \text{ m}. \] (5.36)

**Is this approximate depth an overestimate or underestimate? How accurate is it?**

This approximation neglects the sound-travel time, so it overestimates the free-fall time and therefore the depth. Compared to the true depth of roughly 71.56 m, it overestimates the depth by only 11%—reasonable accuracy for a quick method offering physical insight. Furthermore, this approximation suggests its own refinement.

**How can this approximation be improved?**

To improve it, use the approximate depth \( h_0 \) to approximate the sound-travel time.

\[ t_{\text{sound}} \approx \frac{h_0}{c_s} \approx 0.24 \text{ s}. \] (5.37)

The remaining time is the next approximation to the free-fall time.
\[ t_1 = T - \frac{h_0}{c_s} \approx 3.76 \text{ s}. \] (5.38)

In that time, the rock falls a distance \( gt_1^2/2 \), so the next approximation to the depth is

\[ h_1 = \frac{1}{2} gt_1^2 \approx 70.87 \text{ m}. \] (5.39)

Is this approximate depth an overestimate or underestimate? How accurate is it?

The calculation of \( h_1 \) used \( h_0 \) to estimate the sound-travel time. Because \( h_0 \) overestimates the depth, the procedure overestimates the sound-travel time and, by the same amount, underestimates the free-fall time. Thus \( h_1 \) underestimates the depth. Indeed, \( h_1 \) is slightly smaller than the true depth of roughly 71.56 m—but by only 1.3%.

The method of successive approximation has several advantages over solving the quadratic formula exactly. First, it helps us develop a physical understanding of the system; we realize, for example, that most of the \( T = 4 \text{ s} \) is spent in free fall, so the depth is roughly \( gT^2/2 \). Second, it has a pictorial explanation (Problem 5.34). Third, it gives a sufficiently accurate answer quickly. If you want to know whether it is safe to jump into the well, why calculate the depth to three decimal places?

Finally, the method can handle small changes in the model. Maybe the speed of sound varies with depth, or air resistance becomes important (Problem 5.32). Then the brute-force, quadratic-formula method fails. The quadratic formula and the even messier cubic and the quartic formulas are rare closed-form solutions to complicated equations. Most equations have no closed-form solution. Therefore, a small change to a solvable model usually produces an intractable model—if we demand an exact answer. The method of successive approximation is a robust alternative that produces low-entropy, comprehensible solutions.

**Problem 5.31  Parameter-value inaccuracies**

What is \( h_2 \), the second approximation to the depth? Compare the error in \( h_1 \) and \( h_2 \) with the error made by using \( g = 10 \text{ m s}^{-2} \).

**Problem 5.32  Effect of air resistance**

Roughly what fractional error in the depth is produced by neglecting air resistance (Section 2.4.2)? Compare this error to the error in the first approximation \( h_1 \) and in the second approximation \( h_2 \) (Problem 5.31).
Problem 5.33  Dimensionless form of the well-depth analysis

Even the messiest results are cleaner and have lower entropy in dimensionless form. The four quantities \( h, g, T, \) and \( c_s \) produce two independent dimensionless groups (Section 2.4.1). An intuitively reasonable pair are

\[
\overline{h} \equiv \frac{h}{gT^2} \quad \text{and} \quad \overline{T} \equiv \frac{gT}{c_s}.
\]

(a) What is a physical interpretation of \( \overline{T} \)?

(b) With two groups, the general dimensionless form is \( \overline{h} = f(\overline{T}) \). What is \( \overline{h} \) in the easy case \( \overline{T} \to 0 \)?

(c) Rewrite the quadratic-formula solution

\[
h = \left( \frac{-\sqrt{2/g} + \sqrt{2/g + 4T/c_s}}{2/c_s} \right)^2
\]

as \( \overline{h} = f(\overline{T}) \). Then check that \( f(\overline{T}) \) behaves correctly in the easy case \( \overline{T} \to 0 \).

Problem 5.34  Spacetime diagram of the well depth

How does the spacetime diagram [44] illustrate the successive approximation of the well depth? On the diagram, mark \( h_0 \) (the zeroth approximation to the depth), \( h_1 \), and the exact depth \( h \). Mark \( t_0 \), the zeroth approximation to the free-fall time. Why are portions of the rock and sound-wavefront curves dotted? How would you redraw the diagram if the speed of sound doubled? If \( g \) doubled?

5.5 Daunting trigonometric integral

The final example of taking out the big part is to estimate a daunting trigonometric integral that I learned as an undergraduate. My classmates and I spent many late nights in the physics library solving homework problems; the graduate students, doing the same for their courses, would regale us with their favorite mathematics and physics problems.

The integral appeared on the mathematical-preliminaries exam to enter the Landau Institute for Theoretical Physics in the former USSR. The problem is to evaluate

\[
\int_{-\pi/2}^{\pi/2} (\cos t)^{100} \, dt
\]

(5.42)
to within 5% in less than 5 min without using a calculator or computer! That $(\cos t)^{100}$ looks frightening. Most trigonometric identities do not help. The usually helpful identity $(\cos t)^2 = (\cos 2t - 1)/2$ produces only

$$(\cos t)^{100} = \left(\frac{\cos 2t - 1}{2}\right)^{50},$$

which becomes a trigonometric monster upon expanding the 50th power. A clue pointing to a simpler method is that 5% accuracy is sufficient—so, find the big part! The integrand is largest when $t$ is near zero. There, $\cos t \approx 1 - t^2/2$ (Problem 5.20), so the integrand is roughly

$$(\cos t)^{100} \approx \left(1 - \frac{t^2}{2}\right)^{100}. \tag{5.44}$$

It has the familiar form $(1 + z)^n$, with fractional change $z = -t^2/2$ and exponent $n = 100$. When $t$ is small, $z = -t^2/2$ is tiny, so $(1 + z)^n$ may be approximated using the results of Section 5.3.4:

$$(1 + z)^n \approx \begin{cases} 1 + nz & (z \ll 1 \text{ and } nz \ll 1) \\ e^{nz} & (z \ll 1 \text{ and } nz \text{ unrestricted}). \end{cases} \tag{5.45}$$

Because the exponent $n$ is large, $nz$ can be large even when $t$ and $z$ are small. Therefore, the safest approximation is $(1 + z)^n \approx e^{nz}$; then

$$(\cos t)^{100} \approx \left(1 - \frac{t^2}{2}\right)^{100} \approx e^{-50t^2}. \tag{5.46}$$

A cosine raised to a high power becomes a Gaussian! As a check on this surprising conclusion, computer-generated plots of $(\cos t)^n$ for $n = 1 \ldots 5$ show a Gaussian bell shape taking form as $n$ increases.

Even with this graphical evidence, replacing $(\cos t)^{100}$ by a Gaussian is a bit suspicious. In the original integral, $t$ ranges from $-\pi/2$ to $\pi/2$, and these endpoints are far outside the region where $\cos t \approx 1 - t^2/2$ is an accurate approximation. Fortunately, this issue contributes only a tiny error (Problem 5.35). Ignoring this error turns the original integral into a Gaussian integral with finite limits:

$$\int_{-\pi/2}^{\pi/2} (\cos t)^{100} \, dt \approx \int_{-\pi/2}^{\pi/2} e^{-50t^2} \, dt. \tag{5.47}$$
Unfortunately, with finite limits the integral has no closed form. But extending the limits to infinity produces a closed form while contributing almost no error (Problem 5.36). The approximation chain is now
\[
\int_{-\pi/2}^{\pi/2} (\cos t)^{100} \, dt \approx \int_{-\pi/2}^{\pi/2} e^{-50t^2} \, dt \approx \int_{-\infty}^{\infty} e^{-50t^2} \, dt.
\] (5.48)

**Problem 5.35** Using the original limits
The approximation \(\cos t \approx 1 - t^2/2\) requires that \(t\) be small. Why doesn’t using the approximation outside the small-\(t\) range contribute a significant error?

**Problem 5.36** Extending the limits
Why doesn’t extending the integration limits from \(\pm \pi/2\) to \(\pm \infty\) contribute a significant error?

The last integral is an old friend (Section 2.1): \(\int_{-\infty}^{\infty} e^{-\alpha t^2} \, dt = \sqrt{\pi/\alpha}\). With \(\alpha = 50\), the integral becomes \(\sqrt{\pi/50}\). Conveniently, 50 is roughly 16\(\pi\), so the square root—and our 5\% estimate—is roughly 0.25.

For comparison, the exact integral is (Problem 5.41)
\[
\int_{-\pi/2}^{\pi/2} (\cos t)^n \, dt = 2^{-n} \left( \frac{n}{n/2} \right) \pi.
\] (5.49)

When \(n = 100\), the binomial coefficient and power of two produce
\[
\frac{12611418068195524166851562157}{158456325028528675187087900672} \pi \approx 0.25003696348037.
\] (5.50)

Our 5-minute, within-5\% estimate of 0.25 is accurate to almost 0.01\%!

**Problem 5.37** Sketching the approximations
Plot \((\cos t)^{100}\) and its two approximations \(e^{-50t^2}\) and \(1 - 50t^2\).

**Problem 5.38** Simplest approximation
Use the linear fractional-change approximation \((1 - t^2/2)^{100} \approx 1 - 50t^2\) to approximate the integrand; then integrate it over the range where \(1 - 50t^2\) is positive. How close is the result of this 1-minute method to the exact value 0.2500…?

**Problem 5.39** Huge exponent
Estimate
\[
\int_{-\pi/2}^{\pi/2} (\cos t)^{10000} \, dt.
\] (5.51)
5.6 Summary and further problems

Upon meeting a complicated problem, divide it into a big part—the most important effect—and a correction. Analyze the big part first, and worry about the correction afterward. This successive-approximation approach, a species of divide-and-conquer reasoning, gives results automatically in a low-entropy form. Low-entropy expressions admit few plausible alternatives; they are therefore memorable and comprehensible. In short, approximate results can be more useful than exact results.

Problem 5.42 Large logarithm
What is the big part in \( \ln(1 + e^2) \)? Give a short calculation to estimate \( \ln(1 + e^2) \) to within 2%.

Problem 5.43 Bacterial mutations
In an experiment described in a Caltech biology seminar in the 1990s, researchers repeatedly irradiated a population of bacteria in order to generate mutations. In each round of radiation, 5% of the bacteria got mutated. After 140 rounds, roughly what fraction of bacteria were left unmutated? (The seminar speaker gave the audience 3 s to make a guess, hardly enough time to use or even find a calculator.)
Problem 5.44  Quadratic equations revisited
The following quadratic equation, inspired by [29], describes a very strongly
damped oscillating system.

\[ s^2 + 10^9 s + 1 = 0. \]  \hspace{1cm} (5.54)

a. Use the quadratic formula and a standard calculator to find both roots of the
quadratic. What goes wrong and why?
b. Estimate the roots by taking out the big part. (Hint: Approximate and solve
the equation in appropriate extreme cases.) Then improve the estimates using
successive approximation.
c. What are the advantages and disadvantages of the quadratic-formula analysis
versus successive approximation?

Problem 5.45  Normal approximation to the binomial distribution
The binomial expansion

\[ \left( \frac{1}{2} + \frac{1}{2} \right)^{2n} \]  \hspace{1cm} (5.55)

contains terms of the form

\[ f(k) \equiv \left( \frac{2n}{n-k} \right)^{2-2n}, \]  \hspace{1cm} (5.56)

where \( k = -n \ldots n \). Each term \( f(k) \) is the probability of tossing \( n - k \) heads
(and \( n + k \) tails) in \( 2n \) coin flips; \( f(k) \) is the so-called binomial distribution
with parameters \( p = q = 1/2 \). Approximate this distribution by answering the
following questions:

a. Is \( f(k) \) an even or an odd function of \( k \)? For what \( k \) does \( f(k) \) have its
maximum?
b. Approximate \( f(k) \) when \( k \ll n \) and sketch \( f(k) \). Therefore, derive and explain
the normal approximation to the binomial distribution.
c. Use the normal approximation to show that the variance of this binomial
distribution is \( n/2 \).

Problem 5.46  Beta function
The following integral appears often in Bayesian inference:

\[ f(a, b) = \int_0^1 x^a (1-x)^b \, dx, \]  \hspace{1cm} (5.57)

where \( f(a-1, b-1) \) is the Euler beta function. Use street-fighting methods to
conjecture functional forms for \( f(a,0) \), \( f(a,a) \), and, finally, \( f(a,b) \). Check your
conjectures with a high-quality table of integrals or a computer-algebra system
such as Maxima.