CHAPTER 1
Fundamentals

1.1 Motivation for this book

Hydrology is the study of water, and in the International Glossary of Hydrology (UNESCO/WMO 1992) it is defined as 'Science that deals with the waters above and below the land surfaces of the Earth, their occurrence, circulation and distribution, both in space and time, their biological, chemical and physical properties, their reaction with their environment, including their relation to living beings'. The movement and transformation of water within these processes as described in the definition, as a fluid, will obey the physical rules of fluid mechanics. Fluid mechanics, being a quantitative topic, requires heavy use of mathematical concepts, and these concepts are therefore naturally found in hydrology. These quite basic physical principles can be used effectively to model and hence predict and understand the behaviour of water under many useful circumstances.

Nonetheless, despite the essentially predictable behaviour of water that justifies the use of mathematical principles, often, the flow of water in practice is subject to forces that are beyond our ability to measure with any precision: for example, water in the atmosphere is heated, cooled, mixed with numerous gasses, and transported across large distances under the action of turbulent winds. Eventually, water condenses out of the atmosphere in the form of precipitation but exactly when, where, and how much water falls to the ground under gravity is often extremely uncertain. This uncertainty usually makes it useless to apply the basic physical principles of fluid mechanics to the flow of water in these circumstances. For this reason, hydrologists often turn to statistics, which can be considered as the application of mathematics to uncertain phenomena.

Quantitative hydrology is, therefore, based on an interesting mix of the two great branches of applied mathematics: physical laws (mathematical physics) and probability (mathematical statistics).

Mathematics is, perhaps, the archetypal example of a composite subject. This means that more complex concepts are built from many simpler ones, and so, in order to properly understand the more complex topic, it is necessary to understand the simpler ones from which it is constructed. Not all subjects are like this: it is possible to gain a deep understanding of many aspects of plant biology without

Understanding Mathematical and Statistical Techniques in Hydrology: An Examples-Based Approach, First Edition. Harvey J. E. Rodda and Max A. Little.

having to know anything about mammals, for instance. But mathematics is unforgiving: one cannot understand the true meaning of equations of fluid transport without knowing calculus. Unfortunately, for many reasons, the chance to learn the basic mathematical concepts is not afforded to every student or practitioner of hydrology, and many find themselves at a loss when presented with more complex mathematical concepts as a result.

This book is therefore, intended as a guide to students and practitioners of hydrology without a formal or substantive background in either mathematical physics, or mathematical statistics, who need to gain a more thorough grounding of these mathematical techniques in practical hydrological applications.

1.2 Mathematical preliminaries

This book refers extensively to many essential, but nonetheless quite simple, mathematical concepts; we introduce them here. It is assumed that readers will refer back to this section on reading the later material.

1.2.1 Numbers and operations

Usually when one thinks of 'mathematics', one thinks of numbers, along with operations such as adding, subtracting, multiplying (forming the *product*) and dividing them. Numbers and operations are intimately related: for example, with the simplest of numbers, the *whole numbers*, we can answer questions such as 'what number, when added to 5, gives 10?' Symbolically, we wish to find the x that satisfies the equation $x+5=10$, the answer being $x=5$. But some simple questions involving whole numbers cannot be answered using whole numbers, for instance, the problem 'what number, when added to 10, gives 5?', or $x+10=5$, has no whole number answer. To solve such a problem, we need to include *negative numbers* and *zero*; mathematicians call these whole numbers that can be negative, zero, or positive, the *integers* (all the positive whole numbers are included in the integers). Still, when faced with whole number problems involving multiplication, integers may not suffice. For example, the problem 'what number, when multiplied by 5, equals 1?', or $5x=1$, has no integer solution. The answer $x=1/5$ is called a *rational number* and all the integers are included in the rationals. Finally, it turns out that there are yet more problems involving multiplication that cannot be solved using rationals; consider the problem 'what number, when multiplied by itself, equals 2?' The corresponding equation $x\times x=2$ is solved by the *square root* of 2, $x=\sqrt{2}$, which is an example of a *real number*. The set of real numbers includes all the rationals and numbers such as $\pi=3.14159\ldots$ (which can never be written out to full precision because it has an infinite number of decimal places). With the set of all real numbers, a very large set of problems involving numbers and operations that do actually have a solution can be answered.

It is surprising that even in apparently simple situations such as multiplication and addition with whole numbers, that there are equations that have no solution in the rationals, let alone the integers and whole numbers. Such equations baffled mathematicians until the 19th century when a logically consistent foundation for the real number system was devised. But real numbers do not even suffice for all whole number equations! Consider the equation $x \times x + 2 = 0$; because squaring any number is always positive, there does not seem to be any way to choose a number for x that, when squared, gives a negative number to cancel the 2 and satisfy the equation. Nevertheless, it turns that a consistent solution is possible using *complex numbers*; although abstract, these can be useful in physical problems.

These days, because of their practical utility, real numbers tend to be the lifeblood of quantitative sciences including hydrology. For instance, the average amount of rainfall occurring in one day in one location is often given as a real number in millimetres, to a couple of decimal places where such precision is appropriate. Therefore, most practical problems in hydrology involve solutions that are real numbers given to some limited accuracy appropriate to the problem.

1.2.2 Algebra: rearranging expressions and equations

An important step in the historical development of mathematics was the leap from dealing with specific numbers, to dealing with *any* number by using an abstract symbol to stand for that number (this conceptual leap is usually credited to the great Islamic mathematicians of the medieval period). This is the topic of *algebra*: the study of what happens to these symbols as they are manipulated as if they were numbers. Most quantitative problems in the physical sciences can be expressed and solved algebraically.

Algebra involves very simple rules. Although the rules themselves are elementary, the consequences of those rules can be extremely complex; in fact, much research still goes on today to understand the full, logical consequences of algebra. For this reason, one should not underestimate how difficult it can be to correctly derive the consequences of any particular application of algebra in practice, and it is very much worth the effort to become as familiar as possible with the basic rules.

Today, one usually writes something like x or y when one wants to refer to an abstract number; these are also called *variables* (as opposed to specific numbers, which are *constants*). Then the notation $x + y + 1$ is an *algebraic expression* using these two variables and the constant 1.

Expressions on their own do not 'do' anything; to make expressions useful we need to connect them together into e*quations*, for instance, the equation $x + y + 1 = 0$ states that if the variables x and y are added to the constant 1, then the result must be equal to zero. Alternatively, by manipulating (*rearranging*) this equation, we can get the exactly equivalent statement $x + y = -1$, which is

obtained by subtracting 1 from both sides. This is an example of a basic rule in algebra: in order to rearrange an equation, one has to apply the same operation to both sides of an equation, step by step. This rule ensures that before and after the manipulation, the equation still has the same mathematical meaning.

These *algebraic operations* come in pairs – subtraction is the inverse of addition and division is the inverse of multiplication. What this means, roughly, is that subtraction 'undoes' addition and division 'undoes' multiplication. So, actually, what one is doing when rearranging an equation, is applying a sequence of inverse operation to both sides of an equation.

Rearranging equations is fundamental to the way in which answers to mathematical questions are obtained, often by finding the actual number (value) of some variable. For the equation $x+y=-1$, we only know the value of x and y *implicitly* (through the relationship created between them by the equality). However, it is often difficult (if not impossible) to find the value of x from an implicit equation. In this case, the solution is easy of course: rearrange the equation to find x alone on one side of the equation, for instance, $x=-y-1$ (note that it does not matter on which side x appears). Then, we can usually find a unique value for x, because the right-hand side of the equation is an *explicit* formula for *solving for* the value of x.

The 'art' of rearranging equations to solve for a particular variable, then, is to find a sequence of steps that can be applied to both sides of the equation such that we end up with that variable alone on one side of the equation. Unfortunately there is no general procedure for the 'correct' sequence of steps to apply to any equation: efficient equation solving is often a matter of experience and practice.

The operations of addition and multiplication have the important property that when applied to two or more variables or constants (*terms*), the order in which they are applied does not matter. For instance, for the product $2\times x\times y=2\times y\times x=x\times 2\times y$, etc. The same applies if we replace the product with addition: $2+x+y=2+y+x=x+2+y$. *But* when *combining* different algebraic operations, the order in which variables, constants and operations appear in an expression is critical. For example, $2\times x+y$ is not the same as $2\times(x+y)$. The brackets in the second equation indicate that first, x should be added to y, and then the result should be multiplied by 2. In fact, by *expanding* the brackets, the second expression becomes $2\times x+2\times y$, which makes it clear that it is not the same as $2\times x+y$. A tricky example of this is the expression $x-y$: this is *not* equal to $y-x$. In fact, $x-y=-y+x$. The reason is that actually, the expression $-y$ is shorthand for $(-1)\times y$, and we have to take account of the fact that the multiplication of y by -1 must happen before the addition to x.

The general advice then about rearranging more complex expressions and equations is that carefully and systematically, examine the *order* in which the algebraic operations are supposed to be applied to the terms. Modern algebraic notation has some conventions for this (called *precedence* rules); unless otherwise

overridden using brackets, multiplication and division occur before addition and subtraction. As a case in point, consider the following expression:

$$\frac{x+y}{3y} \tag{1.1}$$

This could be interpreted as follows: first add x to y, then multiply y by 3, and then divide the first result by the second result. It *does not* say, for example, multiply y by 3, divide y by this, and then add x; the following is an example of such kind:

$$x+\frac{y}{3y} \tag{1.2}$$

Another way of explaining the difference is that (1.1) can also be written using brackets as $(x+y)/(3y)$ – then the ordering becomes clear. In (1.1) and (1.2), we can apply some rearrangements that might be useful, for example, by expanding out the 'brackets' in (1.1), we get that $\frac{x+y}{3y}=\frac{x}{3y}+\frac{y}{3y}$ (applying the rule that dividing by some expression is equivalent to multiplying by 1 divided by that whole expression). Next, we can apply the rule that dividing an expression by itself is equal to 1 (unless that expression is equal to zero – see below); so $\frac{x+y}{3y}=\frac{x}{3y}+\frac{1}{3}$. For (1.2), we get $x+\frac{y}{3y}=x+\frac{1}{3}$ for the same reason. Slightly more complex is the situation where the top and bottom expressions both involve addition, for instance, as follows:

$$\frac{x+y}{3x+3y} \tag{1.3}$$

To rearrange this expression, we consider *factoring* the bottom part of the division as a rearrangement step. A *factor* is a number (or variable) that multiplies another expression, for instance, the expression $6xy$ has the factors 6, x and y (actually, since $6=2\times 3$, it is also reasonable to argue that there are four factors 2, 3, x and y). The factored bottom expression is then $3x+3y=3(x+y)$. Effectively, we have changed the order of the multiplication by 3 and the addition: this is what factoring achieves. So, factoring undoes expanding out brackets. Now it is clear to see that if $x+y$ is not zero, (1.3) has the value 1/3, that is the $x+y$ expression cancels completely.

The number zero has a special importance in algebra. Firstly, note that adding zero to some expression leaves that expression unchanged, for example $x+0=0+x=x$ (An interesting observation is that 1 plays the same role in multiplication as 0 takes in addition, namely, it leaves the expression unchanged: $x\times 1=1x=x$.) The second property of zero is that multiplying some expression by zero results in zero, for example $0x=x\times 0=0$. Sometimes, we end up with an equation such as $0=xy$. Using the second property, we can see that one or more of x and y must be zero for this equation to be satisfied. Another consequence of these properties is that dividing anything by zero is undefined (effectively, there is

no meaningful result). Consider what it means to write $x = y/0$. For the moment, treating 0 as a symbol, we could rearrange this equation to $0x = y$, and applying the second property of zero y must be zero. However, then the equation becomes $0x = 0$, and this is true for *any* value of x! In other words, the original equation, even though it is an explicit formula for x, does not tell us what value x should take. Because of this, (1.1) and (1.2) are meaningless in the special case where $y = 0$ and (1.3) is meaningless if $x + y = 0$.

Often, we have the situation where there are two or more variables whose value we need to find in order to solve a practical problem. In general, we need as many equations as there are variables in order to find a solution that gives the unique values of all the variables. For instance, to solve the following pair of equations for x and y,

$$x + y = 0 \tag{1.4a}$$

$$3x + y = 1 \tag{1.4b}$$

We might want to solve for x in (1.4a), $x = -y$, and then *substitute* this expression for x into (1.4b), $3(-y) + y = 1$. We can then factor out y in this to obtain $(-3 + 1)y = 1$, and we get an explicit formula $y = 1/(-2) = -1/2$. Now, using the explicit formula for x, it must be that $x = 1/2$, and we have a solution for both variables. This is a simple example that illustrates how to apply sequential rearrangement and substitution in order to solve a pair of equations; this basic principle can be attempted for more complex equations but it usually becomes very difficult in practice to solve equations involving three or more variables. Typically, one then turns to *computer algebra* software, or, instead uses *numerical methods* to obtain approximate solutions.

Repeated self-multiplication of some term or expression has a special name: *exponentiation* ('raising to the power') and is written using the superscript notation as x^n, where n is called the *power* or *exponent*. If n is a whole number, this just means that we multiply x by itself n times. So, that means that $x^1 = x$. If we have powers n and m that are both whole numbers, then it is fairly easy to see that $x^n \times x^m = x^{n+m}$. In a sense, we can see that this rule 'converts' multiplication into addition. When applied to expressions, there are some simple consequences, for instance when $n = 2$, the expression $(x + y)^2 = x^2 + 2xy + y^2$ (which one can check by expanding out the brackets – there is a general formula called the *binomial expansion* that works for any general value of n).

If we allow $n = 0$, then $x^0 \times x^m = x^{0+m} = x^m = 1 \times x^m$, so it is reasonable to claim that $x^0 = 1$ (to be consistent with the role that 1 plays in multiplication, as discussed above). Similarly, if we allow that $n = -m$, then we get $x^{-m} \times x^m = x^{m-m} = x^0 = 1$, so we can claim that $x^{-m} = 1/(x^m)$ to be consistent with the idea that dividing some expression by itself is equal to 1. It follows then that $x^{-1} = 1/x$. In fact, it can be shown that this rule converting multiplication into addition works quite generally: n and m can be any fraction or real number and thus exponentiation is a

general algebraic operation. Certain rational powers are given special names: $x^{1/2}$ is called the square root and written $\sqrt{x}$, and more generally, $x^{1/n}$ is the nth *root* written $\sqrt[n]{x}$.

Being a general algebraic operation, exponentiation has an inverse called the *logarithm* (a key mathematical discovery credited to John Napier in the 16th century). We write this as $\log_x y$, where x is called the *base* of the logarithm, which has the meaning that if $n = \log_x y$, then $x^n = y$: the logarithm to base x recovers the power of x (so, for example, $\log_x \sqrt{x} = 1/2$). As we demonstrated above, since exponentiation 'converts' multiplication into addition, we can explain that the logarithm converts addition back to multiplication. Consider the power rule $x^n \times x^m = x^{n+m}$, then taking the logarithm to base x on both sides, we get $\log_x(x^n \times x^m) = \log_x(x^{n+m}) = n + m = \log_x x^n + \log_x x^m$. As with exponentiation, this rule actually works for general numbers, not just whole numbers, and we can derive some consequences worth memorizing: $\log_x 1 = 0$ (which is the inverse of $x^0 = 1$), $\log_x x = 1$ (which is the inverse of $x^1 = x$), $\log_x x^n = n$, and the general rule $\log_x(a^n \times b^m) = n\log_x a + m\log_x b$ for any numbers a, b, n, m and x, provided only that neither a nor b is zero. The last rule has a useful special case: $\log_x(a/b) = \log_x(a \times b^{-1}) = \log_x a - \log_x b$.

In practice, since logarithms in one base can be converted to any other base using the formula $\log_x a = \log_y a / \log_y x$, one tends to work in a standard base such as 10. The other commonly used base is the *natural logarithm* which uses the base $e = 2.71828\ldots$ (we will see later that this has a very important origin), written as $\ln x$. The inverse to the natural logarithm, $e^x = \exp(x)$, plays a very important role in much of mathematics: as the inverse it follows that $\ln(\exp(x)) = x$.

1.2.3 Functions

Expressions are very often 'packaged up' into convenient shorthand notation known as *functions*, such as $f(x)$ or $g(x)$. Examples of functions include the exponential function $\exp(x) = e^x$ and $\ln(x)$ above but also familiar functions such as the trigonometric functions $\sin(x)$, $\cos(x)$ and $\tan(x)$. Use of functions in expressions can improve the readability of equations considerably. Very often there is an associated *inverse function*: as we have seen, $\exp(x)$ has $\ln(x)$ as its inverse. Sometimes, consideration of the range of acceptable values that a function can take tells us about the range of output values of its inverse: for example, the $\sin(x)$ function takes all possible real *angles* as input, but its output is restricted to the range −1 to 1. So, the inverse function $\sin^{-1} x$ can only *accept* numbers in the range −1 to 1.

Plotting a function as a *graph* can be very useful; typically this is done by drawing a curve on axes where x is on the horizontal and $y = f(x)$ is on the vertical. Then, since a function only outputs one value per unique input value, the curve must be a single, non-self-intersecting line. In addition, very often that line can be drawn without taking the 'pen' off the paper, so the function has no *discontinuities*. Functions can take more than one number as input, for

example $f(x,y) = x^2 + 3y^3$; this makes it much harder to plot a graph of the function (which would appear as a surface in 3D with $z = f(x,y)$ being the height of the surface).

1.2.4 Calculus

The name given to the theory of mathematics that deals with the abstract concepts of area (including length and volume more generally) and gradient (slope) is called *calculus*. Although the mathematicians of the ancient world knew how to calculate these quantities for simple shapes (for example, working out how to divide up a rectangular field into equal areas for the purpose of probate law), they did not know how to do this for general geometric objects, particularly if they had arbitrarily curved boundaries. This had to wait until the 17th century for the mathematical innovations of Newton and Leibniz, who saw the potential for applying and extending these concepts to predicting the motion of the planets. It is probably fair to say that the vast majority of physical applied maths revolves around the use of concepts from calculus.

Summation plays a central role in calculus: we write $\sum_{i=0}^{N} x_i$ to denote the sum from 0 to N of the values in the $N+1$ variables x_i (we use the subscript notation to index each of these different variables). We can apply this to the problem of calculating areas – the area of a rectangle is just $A = w \times h$ where w is the width and h is the height. Now, if a given shape can be approximately broken down into $N+1$ small rectangles, then the area of the complete shape is approximately as follows:

$$A \approx \sum_{i=0}^{N} w_i h_i \tag{1.5}$$

In words, 'the sum of the product of the width of each rectangle times the height of rectangle is approximately the total area of the shape'.

If we can assume that the width of all these rectangles is the same, we can simplify this to $\sum_{i=0}^{N} w h_i$. Of course, this will only be an approximation to the area, for example some of the area might not be counted.

Assuming, for the sake of simplicity, that one edge of the object is straight and lies on the x-axis of graph and the other side is represented by a function with arbitrary curves (this requirement might seem contrived but it turns out that calculus can be defined in more flexible ways for different geometric situations, using essentially the same ideas). Figure 1.1 shows this idea for finding the area (*integrating*) under the curve $f(x) = x^2$. In the upper panel, we have a relatively coarse set of rectangles with equal width w attempting to fill the area; the bottom panel has a much slimmer set of rectangles, again of fixed width. It is easy to see that the amount of uncounted area in the bottom panel is smaller than that in the top panel, so the bottom panel is a better approximation to the area. In calculus,

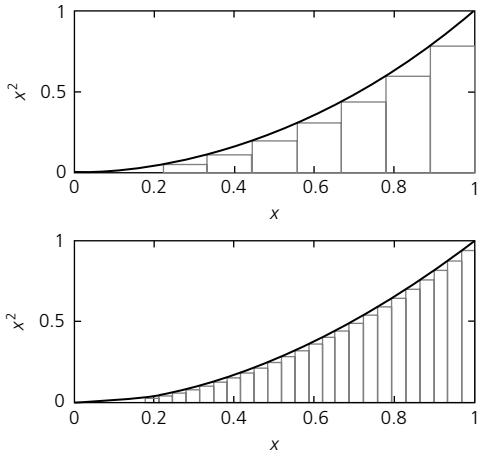

Figure 1.1 Approximate integration of the area under the curve x^2 (black) using rectangles (grey), over the interval 0–1, with coarse partition (top) and finer partition (bottom).

the idea is to calculate what happens as the width of the rectangles becomes arbitrarily small, following from the intuition that slimmer rectangles give better approximations: of course, the number of rectangles will become arbitrarily large as a result. The approach aims to convert the problem of finding the area under the curve to a problem of finding the ultimate value of a sequence of better and better approximations.

To do this, we will need the idea of *limits*, one of the core concepts of calculus. Mathematicians use the shorthand notation '$a = \lim_{x \to c} f(x)$' for the limiting value of the function $a = f(x)$ as x takes on values that are always getting closer to c. This is also written as '$f(x) \to a$ as $x \to c$'. A critical point to understand is that in most useful cases, the limiting value a cannot be calculated directly. For example, it makes intuitive sense (and it is logically correct as we will show next) that $\lim_{x \to \infty} 1/x = 0$. But since infinity does not have a *definite* value, algebraic expressions such as $1/\infty$ do not have a definite result either.

Limits are a (indirect) way of computing definite answers in these situations. For example, we know that the function $f(x) = 1/x$ is continuous (see above) at all values of x except 0. Also, the function is *decreasing*, that is if we pick any two numbers x and y such that $x < y$, then $1/x > 1/y$. Additionally, we know that if x is positive, then $1/x$ is also positive. These pieces of information allow us to conclude that $1/x \to 0$ as $x \to \infty$. In other words, we have shown that as we keep increasing the (positive) value of x, $1/x$ always gets smaller, and since it cannot be negative, as x becomes arbitrarily large (infinite), $1/x$ must ultimately take on the value zero. At root, this is typical of limit value arguments: nonetheless, most problems

encountered in practice are reducible to an algebraic combination of known results about the limits of basic functions.

We now return to the problem of finding the limit of sequence of approximations to the area under the curve. We can construct a grid of x-values as $x_i = wi$ and the corresponding height of the rectangles is $h_i = (wi)^2$. Then the number of rectangles in the interval 0–1 is $N = 1/w$. So, the area is written as:

$$A = \lim_{w \to 0} \sum_{i=0}^{N} w(wi)^2 \tag{1.6}$$

This equation states that first, we sum up all the areas of the rectangles fitting underneath the curve. Then, we take the limit of these sums, as the width of these rectangles becomes arbitrarily small. For every rectangle width, there will be a corresponding number of rectangles N, which therefore must go to infinity as the width goes to zero.

Here we make the remark that in this specific case (1.6) does have an exact answer, $A = 1/3$, that we can compute using well-known, but somewhat complex, algebraic manipulations.

If we want to compute the area under an arbitrary function $f(x)$ over any chosen range of values of the x-axis, say, from a to b, we need the *definite integral*:

$$\int_a^b f(x)dx = \lim_{N \to \infty} \sum_{i=0}^{N} f(a + iw)w \tag{1.7}$$

where we choose $N = (b-a)/w$ (note the w is often written as Δx, because w is a 'small difference in x'). In Equation (1.7), the left-hand side is just shorthand for the right-hand side, which states that the area is computed by summing up rectangles of width w, placed at each position on a grid of spacing w covering from a and b on the x-axis. The particular choice of N means that when $i = N$, $a + iw = b$, the right-most grid position. Each rectangle has height $f(a + iw)$.

Note this is only a definition: there is no guarantee that we can actually find the limit of the sequence of approximations to find the exact answer by some straightforward algebra. In fact, the somewhat disappointing news is that the number of functions $f(x)$ that we *cannot* integrate in this way vastly outnumbers the functions that can be integrated like (1.6). This usually happens because the kind of algebraic tricks used to remove the summation in situations such as (1.6) work only in special cases. Nonetheless, under certain conditions that are not too restrictive, we can say that the limit in (1.7) is useful, in that, it has a definite value, and we can approximate this to any desired accuracy using a computer program, for example.

The inverse operation to integration is *differentiation*. It is relatively simple to find the gradient of a straight line. Imagine finding the slope of a straight road running up a hill with constant angle to the horizontal: it is just the *rise over the run*

or the change in vertical height (Δy) you go through as you travel over some horizontal distance (Δx):

$$m = \frac{\Delta y}{\Delta x} \tag{1.8}$$

This is how to calculate the gradient of a function if it is a straight line. How can we do this if the function is not a straight line? One way is to assume that over small enough distances, the slope of *any* function at a particular fixed point can be approximated by the slope of a straight line that goes through that point. This will be a good assumption if the function is smooth enough. As the distance over which we make this assumption gets smaller, the approximation to the slope at that point gets better. For a given function $f(x)$, the change in 'height' at x over the distance Δx is $f(x+\Delta x)-f(x)$, and therefore, using the idea of limits to define the derivative, we get:

$$m(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x} \tag{1.9}$$

(Note that the slope of a general function is itself a function of x, the chosen point, unlike a straight line, which has the same slope at every point.) In this way, *differentiation* solves the problem of how to find the slope of a function which is arbitrarily curvy. The derivative is also commonly written as df/dx or also $f'(x)$ when it is clear that we are differentiating with respect to x.

Algebra that arises from differentiating is usually a lot simpler than algebra that arises as a result of integration. For this reason, many more functions can be algebraically differentiated than integrated. As with integration, under not very restrictive conditions, the limit in (1.9) is useful and can be calculated approximately to any desired precision numerically.

There is a theorem in calculus that relates integration and differentiation (called, appropriately enough, the 'fundamental theorem of calculus'). This theorem can be stated in many different ways, but it is instructive to provide geometric intuition. Firstly, consider the area under the function $f(x)$ from 0 to x written as $F(x)$. Now the area under the curve between x and $x+\Delta x$ can be found as $F(x+\Delta x)-F(x)$, which is the area from 0 to $x+\Delta x$ minus the area from 0 to x. However, as above, when defining the integral, we could also approximate the area between x and $x+\Delta x$ with the small rectangle of area $f(x)\Delta x$. It follows that $F(x+\Delta x)-F(x) \approx f(x)\Delta x$ or

$$f(x) \approx \frac{F(x+\Delta x)-F(x)}{\Delta x} \tag{1.10}$$

Taking the limit of both sides as $\Delta x \to 0$ gives us $f(x) = F'(x)$, using the definition of the derivative above. So, this says that the original function $f(x)$ is what we get by finding the slope of the area under the curve of that function.

The following is the most common statement of the fundamental theorem:

$$\int_a^b f(x)dx = F(b) - F(a) \tag{1.11}$$

This says that the definite integral of the function $f(x)$ is the area under the function from 0 to b minus the area under the function from 0 to a. This allows us to introduce the *indefinite integral*, useful when integrating from 0 to the value of some variable x:

$$\int f(x)dx = \int_0^x f(x')dx' = F(x) - F(0) = F(x) + c \tag{1.12}$$

(Note that we had to use a *dummy variable* x' to avoid a conflict between variable names, because x normally does not appear in the integration range). The replacement of $-F(0)$ with the generic constant c indicates that an arbitrary constant is always introduced when performing indefinite integration.

Yet another statement of the fundamental theorem is:

$$\int_a^b f'(x)dx = f(b) - f(a) \tag{1.13}$$

We get this from (1.11) by replacing the function $f(x)$ with the derivative of the function $f'(x)$ instead. This form of the theorem tells us something quite profound about calculus that has far-reaching consequences in many areas of mathematics: the definite integral of the slope of the function is just the difference of the function value at the far end of the range, minus the function value at the near end.

The derivative and integral have certain important algebraic properties of their own that are consequences of the way they are defined. The first is the fact that they are both *linear operations*:

$$\int [af(x) + bg(x)]dx = a\int f(x)dx + b\int g(x)dx \tag{1.14a}$$

$$\frac{d}{dx}[af(x) + bg(x)] = a\frac{df}{dx}(x) + b\frac{dg}{dx}(x) \tag{1.14b}$$

In the above, $f(x)$ and $g(x)$ are arbitrary functions, and a, b are constants, and the equations state that it is possible to swap the order of scaling, addition and integration, and the same with differentiation. This means that we can first multiply two functions by constants, add the results together and then integrate, or we can first integrate the functions separately, multiply the results by constants and then add the results together. The same applies to differentiation. It is critical to note that the above rules only apply if a, b do not change as x changes.

More algebraic properties apply to differentiation, for example, the *product rule* states what happens when differentiating the product of two functions:

$$\frac{d}{dx}f(x)g(x) = f(x)\frac{dg}{dx}(x) + g(x)\frac{df}{dx}(x) \tag{1.15}$$

Similarly, the *chain rule* explains what happens if we differentiate a function of a function:

$$\frac{d}{dx}f(g(x)) = \frac{df}{dx}(g(x))\frac{dg}{dx}(x) \tag{1.16}$$

There are further 'rules' that occur in more complex combinations. It is important to grasp that (1.14a) is the *only* genuine algebraic property of differentiation shared by integration – properties (1.15), (1.16) have no direct counterparts in integration. There are other integration 'rules', such as the integration by parts and integration by substitution, but these are actually obtained by 'undoing' the rules of differentiation (1.15) and (1.16).

Differentiating a function twice gives the *second derivative* (also known as the *curvature* of a function):

$$\frac{d}{dx}\left(\frac{d}{dx}f(x)\right) = \frac{d^2}{dx^2}f(x) = f''(x) \tag{1.17}$$

Similarly, the nth *derivative* is written as d^n/dx^n, for $n > 0$, or sometimes $f^{(n)}(x)$.

It is helpful to list a few specific derivatives and integrals. Perhaps the most important is the exponential function encountered earlier:

$$\frac{d}{dx}\exp(x) = \exp(x) \tag{1.18a}$$

$$\int \exp(x)dx = \exp(x) + c \tag{1.18b}$$

This explains one important reason behind the special place of the exponential function in mathematics as it is the only function that is simultaneously its own integral and derivative. Other important functions include the powers of x (*polynomials in x*):

$$\frac{d}{dx}x^n = nx^{n-1} \tag{1.19a}$$

$$\int x^n dx = \frac{1}{(n+1)}x^{n+1} + c \tag{1.19b}$$

Many other functions are explicitly differentiable and integrable in this way, notably the trigonometric functions (sine, cosine, tangent), but most functions do not have simple integrals and we usually turn to numerical algorithms to compute them for particular ranges in practice.

1.2.5 Differential equations

Having defined differentiation as the slope of a function with arbitrary 'curviness', since the slope depends on the chosen value of x, the derivative of a function is a new function of x. So, as with any function, it can usefully appear in an equation alongside other functions, operations, and constants. The resulting *differential equations* have been the cornerstone of modern physical applied mathematics ever since Newton's *Principia*: the sheer number of physical problems that can be formalized using differential equations is truly staggering.

As an example, consider the following differential equation:

$$\frac{d}{dx}f(x) = m \tag{1.20}$$

where m is just a constant that we know. This states that the equation is satisfied if $f(x)$, when differentiated, is constant. A moment's thought will lead to the answer that $f(x)$ must describe a straight line on the graph of the function: there is no other function whose slope is always the same constant. We can also prove this by integrating both sides of (1.20), finding that $f(x) = mx + c$, which is indeed an expression for a line on the graph of x against $f(x)$, called the *general solution* to the differential Equation (1.20). We can check that we have the right solution by differentiating $f(x)$, and testing that this satisfies the equation. One important point to note is that an arbitrary constant c appears in the solution (because we are performing indefinite integration) so, without specifying this constant, we cannot find the value of the solution for a given x. In this case, the constant can be set by specifying an *initial condition*: that is what we expect $f(x)$ to be when $x = 0$. For instance, the initial condition $f(0) = -2$ leads to the *specific solution* $f(x) = mx - 2$.

Perhaps the most famous of all elementary differential equations is the simplified model of the mass on a spring, ignoring friction:

$$mf''(t) = -kf(t) \tag{1.21}$$

where m is the mass, and k is the spring stiffness, and t represents time. In physics, the quantities $f(t)$, $f'(t)$ and $f''(t)$ have special names: *position*, *velocity*, and *acceleration* (those with a physics background might recognise (1.21) as an application of *Newton's second law*). So, (1.21) states that the acceleration, multiplied by the mass, is equal to the negative of the position multiplied by the spring constant. We will also assume that the position at time zero is some constant A: $f(0) = A$, and the initial velocity is zero: $f'(0) = 0$. Special algebraic techniques have been developed to find solutions to equations such as (1.21), when applied, the specific solution is:

$$f(t) = A\cos\left(\sqrt{\frac{k}{m}}t\right) \tag{1.22}$$

The cosine arises because it can be shown that if $f(x) = \cos(x)$, then $f''(x) = -\cos(x)$, that is the cosine function is the negative of its' own second derivative, which is, essentially, what is required to satisfy (1.21). The character of this solution is *oscillatory*: that is, the mass vibrates at a rate of $\sqrt{k/m}$, with amplitude A. So, if the mass is increased, the vibration becomes slower, and if the spring stiffness is increased, the vibration speeds up (an intuitive result). Note that the rate of vibration is not dependent upon the initial position A.

Differential equations used in quantitative hydrology can be a lot more complex than (1.21), but the principles are the same. Most of the complexity arises when dealing with functions of more than one variable. For example, a function $Q(x, t)$ might represent the quantity of water in a channel in both time and position. Then, we need to introduce *partial differential equations* that involve the derivative of the function in one or more variable at a time, for example, the (one-way) *kinematic wave equation* in hydrological flow routing is:

$$c\frac{\partial Q}{\partial x}(x,t) = -\frac{\partial Q}{\partial t}(x,t) \tag{1.23}$$

The notation $\partial Q/\partial x$ is shorthand for the derivative of the function $Q(x, t)$ with respect to x alone:

$$\frac{\partial Q}{\partial x}(x,t) = \lim_{\Delta x \to 0} \frac{Q(x+\Delta x,t) - Q(x,t)}{\Delta x} \tag{1.24}$$

Equation (1.23) states that the rate of change of the quantity in space, multiplied by c is the negative of the rate of change of the quantity in time. Again, techniques have been developed to solve such equations to find an explicit expression for $Q(x, t)$. This equation has some similarities to (1.21): except that it involves only first derivatives, and two variables instead of one.

Many of the equations of quantitative hydrology are partial differential equations such as (1.23). A large number of useful ones (such as the *shallow water wave equation*) are, unfortunately, unsolvable using the kind of algebraic tricks that are available for solving (1.21) and (1.23). For this reason, computational algorithms involving purely numerical calculations have been devised and are an important tool in modern quantitative hydrology.

1.2.6 Probability and statistics

Statistical techniques form a critical part of the material in Chapters 2, 6 and 7. Statistics is based on the mathematics of uncertainty, known as *probability*. By comparison to the other areas of mathematics covered above, probability as a mathematical topic is a relative newcomer, having origins in the 17th century. As with algebra, the basic rules of probability are elementary and intuitive: but the logical consequences of these rules, particularly when applied to real-world data, can be complex and often counter-intuitive. For instance, consider the

chance of rain falling on the city of Oxford in the UK in any one day. The (extensive) historical data suggests that it is as likely to rain on any day as not, and that whether it rained yesterday or the day before has almost no influence on whether it rains today. So it is quite accurate to model this situation as the flipping of an unbiased coin: the probability of rainfall in one day is 1/2, and previous coin flips do not influence the current one. When this is explained to people, most will surmise, correctly (presumably from experience), that the probability of there being 30 rain days in a row is extremely small. Because of this, if, in the unlikely event that 30 days of rain did actually occur, many will assume that the next 30 days will have to be dry in order to maintain the expected probability of 1/2. But this belief is false: according to the coin-flipping model, whether it rained yesterday or on any previous day has no bearing on whether it rains today. We have simply witnessed an extraordinarily unusual event. Of course, if such an event *did* occur in the historical record, it would distort the statistics so much that we might decide that the unbiased coin model is not actually appropriate.

The mathematical ingredients of probability are easy to state. There is the set of all possible *outcomes* relevant to the physical situation. For example, the cloudiness at any one time in one location can be observed as clear, scattered clouds (~25%), partly cloudy (50% coverage), mostly cloudy (75% covered) or overcast; or it can rain or not on any one day at a specific location; or the rainfall in any one day in one location can be zero or more millimetres, in steps of one-tenth of a millimetre. From these outcomes, we form *events* of interest, the probability of which we want to know. For example, the event that the rainfall depth is greater than or equal to 10.0 mm, or the event that there are scattered clouds.

To these basic ingredients are added three rules (known as *axioms* by mathematicians):

Rule 1. To each possible event is attached a real number called the *probability value* that must not be negative;

Rule 2. Since one of the outcomes is certain to occur, the event that any one of the outcomes is observed is assigned the probability value 1. From this, and the previous condition, we can conclude that probabilities of events lie between 0 and 1 inclusive, with 0 meaning *impossible*, and 1 denoting *inevitable*;

Rule 3. The probability of any compound event obtained by joining mutually exclusive (that is, non-overlapping) events together, is just the sum of the probabilities of the individual events.

Some examples are useful. Consider the event that rainfall is at least 10.0 mm, and assume that the probability of this event is known to be 0.01. The *complementary* event is that the rainfall is less than 10.0 mm. These two events are mutually exclusive because any outcome (rainfall depth in 0.1 mm steps) satisfies one of the

events, but not both simultaneously. In fact, these two events are also *exhaustive* because joined together, they cover any outcome. The probability of the 'joint' event, which is just 'some rainfall depth occurs' is 1 (from rule 2). This means that the probability of the complement event (rainfall depth is <10.0) must be $1-0.01=0.99$ (to satisfy rule 3).

As another simple example: if we know that any of the five categories of cloud cover are equally likely, then the probability of any category is just 1/5. Then, since it cannot be both clear and overcast at the same time, the probability of the location being overcast or clear is $1/5+1/5=2/5$ (using rule 3).

Usually, when dealing with random quantities in physical problems, there is an intuitive numerical label for each possible outcome. If this label is a whole number or integer, then the random quantity X can be associated with a *probability mass function* P (PMF – also known as a *distribution*), that determines the probability value assigned to each event, which lies between 0 and 1. From the rules above, it must be the case that the sum of the distribution of each outcome must be 1. For example, if the outcomes are labelled as integers from zero and above, then $\sum_{i=0}^{\infty} P(X=i)=1$ to satisfy rule 2.

When the random quantity is a real number, say, depth of a river, then the quantity X is associated with a *probability density function* p (PDF – although it is common to call this a distribution as well). With real-valued variables, some subtleties occur. Firstly, the PDF itself does not represent a probability value: we need to invoke calculus to find the *area under the PDF* which gives the probability. For example, the probability that X lies between 2 and 3, for instance, is $\int_2^3 p(x)dx$. From the property of the integral that $\int_a^a p(x)dx=0$, we can infer that the probability that X takes on some single value a is always actually zero. Also, the probability density function can be larger than 1: this does not violate rule 2, because the area under the PDF must sum to 1, that is $\int p(x)dx=1$. Therefore, the area under any smaller range of values than all possible values of X will be less than 1.

Statistical hydrology is often interested in calculating the probability of some event based on a *probability density model* for that variable. Many of the models have *parameters*: constants that affect the shape of the density function. Somehow, these parameters have to *estimated* for a particular data record. Parameter estimation is one of the main topics of statistics, which has led to a large range of techniques. Perhaps the most widely-used technique is *maximum likelihood*, which proposes that the optimum parameter values are those that maximize the probability density given the data record.

Particular quantities of distributions have special significance in probability and statistics. The *mean* (often called the *average*) is the *first moment* often written $E[X]$ or $\bar{x}$, and the *variance*, which is the *second (central) moment* written as $E\left[(X-E[X])^2\right]$ or $\mathrm{var}(X)$. Note that the standard deviation is the square root of the variance.

If we know a model distribution for the random variable X, say, $p(x)$, then these quantities can be calculated using integration:

$$\bar{x} = E[X] = \int x\, p(x) dx \tag{1.25}$$

$$\mathrm{var}(X) = E\left[(X - E[X])^2\right] = \int (x - \bar{x})^2\, p(x) dx \tag{1.26}$$

We can also estimate these quantities from data. Technically, this is usually done by constructing a distribution based on the data, which places equal weight at each of the N data points $x_1, x_2 \ldots x_N$, and having no density anywhere else. Then, the integrals in Equations (1.25) and (1.26) simplify to:

$$\bar{x} = E[X] = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1.27}$$

$$\mathrm{var}(X) = E\left[(X - E[X])^2\right] = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{1.28}$$

When calculating these values from data, the accuracy of the estimates is very much tied to the amount of data available: however, in most cases it can be shown that the estimated mean improves as the length of the data record improves.

Finally, another important value is the *median*, which is the special value a such that the probability of X being less than a is the same as the probability of it being larger than a: it can be estimated as the value that lies 'in the middle' when all the data is sorted numerically. Similarly, the maximum is the largest value when the data is sorted.

Reference

UNESCO/WMO (1992) *International Glossary of Hydrology*, 413pp. UNESCO/WMO, Paris/Geneva.