APPENDIX B

MATHEMATICAL FORMALITIES AND STYLE

We collect here a brief account of the various mathematical conventions used throughout this work, and the reasons for them. Careless notation – inadequate, inconsistent, or ambiguous – has led to so many erroneous results in the recent literature that we need to find rules of notation and terminology that make it as difficult as possible to commit such errors.

A mathematical notation, like a language, is not an end in itself but only a communication device, a means for conveying certain ideas. Its purpose is best served if the notation, like the language, is allowed to evolve with use. This evolution usually takes the form of abbreviations for whatever expressions recur often, and reducing the number of symbols when their meaning can be read from the context.

But a living, changing language still needs a kind of safe harbor in the form of a fixed set of rules of grammar and orthography, hidden away in a Dictionary for use when ambiguities threaten. Likewise, probability theory needs a fixed set of normative rules on which we can fall back in case of doubt. We state here our formal rules of notation and logical hierarchy; all Chapters from #3 on start with these standard forms, and evolve from them. A notation which is so convenient that it is almost a necessity in one Chapter, might be only confusing in the next; so each separate topic must be allowed its own independent evolution from the standard beginning.

Notation and Logical Hierarchy

In our formal probability symbols (those with a capital P)

$$P(A|B) \tag{B–1}$$

the entries A, B always stand for *propositions*, with a sufficiently clear meaning (at least to us) that we are willing to use them as elements of Aristotelian logic, obeying a Boolean algebra. Thus $P(A|B)$ does not denote a "function" in the usual sense.

We repeat the warning that a probability symbol is undefined and meaningless if the conditioning statement B happens to have zero probability in the context of our problem [for example, if $B = CD$, but $P(C|D) = 0$]. Failure to recognize this can lead to erroneous calculations – just as, in elementary algebra, inadvertently dividing by an expression that happens to have the value zero, can invalidate all subsequent results.

To preserve the purity of our probability symbols (B–1) we must have also other symbols for probabilities. Thus, if proposition A has the meaning

$$A \equiv \text{“The variable } q \text{ has the particular value } q' \tag{B–2}$$

there is a tendency to write, instead of $P(A|B)$,

$$P(q'|B) \tag{B–3}$$

But q' is not a proposition, and so the writer evidently intends the symbol (B–3) to stand now for an ordinary mathematical function of the variable q'. In our system this is illegitimate, and so when an ordinary mathematical function is intended, we shall take the precaution of inventing a different functional symbol such as $f(\quad|\quad)$, writing (B–3) instead as

$$f(q'|B) \tag{B–4}$$

Now the distinction between (B–3) and (B–4) may appear to some readers as pedantic nit–picking; so why do we insist on it? Thirty years ago, the present writer would also have dismissed this point as too trivial to deserve mention; but later experience has brought to light cases where failure to maintain the distinction in clear sight has tricked writers into erroneous calculations and conclusions. The amount of time and effort this has wasted – and which is still being wasted in this field – justifies our taking protective measures against it.

The point is that a proposition A is a verbal statement that may indeed specify the value of some variable q; but it generally contains qualifying clauses also:

$$A \equiv \text{"The variable } q \text{ has the value } q' \textit{ if the proposition } I \textit{ is true"} \tag{B–5}$$

If we try to take the short–cut of replacing A by q' in the probability symbol, we lose sight of the qualification. Later in the calculation, the same variable q may appear in a proposition A_1 with a different qualification; and again one may be tempted to replace A_1 by q' in the probability symbol. Still later in the calculation the same probability symbol will appear with two different meanings, and one is tricked into supposing that they represent the same quantity.

This is what happened in the famous "marginalization paradox", in which the same probability symbol was used to denote probabilities conditional on two different pieces of prior information, with bizarre consequences described in Jaynes (1980) and in Chapter 15. The confusion is still causing trouble in probability theory, for those who have not yet understood it. The notation (B–4) cannot make such errors impossible, but it makes them less likely, because when the user must invent a new functional symbol, he will probably have the good sense to invent a different one when a different qualification is present.

However, we are not fanatics about this. In cases so simple that there is very little danger of error anyway, we allow a compromise and follow the custom of most writers even though it is not a strictly consistent notation. In probability symbols with a small p, we shall allow the arguments to be either propositions or numbers: thus if A is a proposition and q a number, the equation

$$p(A|B) = p(q|B)$$

is permitted; but with the warning that when small p symbols are used, the reader must judge their meaning from the context, and there is a possibility of error from failure to read them correctly.

A common and useful custom is to use Greek letters to denote continuously variable parameters, Latin letters for discrete indices or data values. We shall adhere to this except when it would conflict strongly with another long established usage.

Our "Cautious Approach" Policy

The derivation of the rules of probability theory from simple desiderata of rationality and consistency in Chapter 2 applied to discrete, finite sets of propositions. Finite sets are therefore our safe harbor, where Cox's theorems apply and nobody has ever been able to produce an inconsistency from application of the sum and product rules. Likewise, in elementary arithmetic finite sets are the safe harbor in which nobody has been able to produce an inconsistency from applying the rules of addition and multiplication.

But as soon as we try to extend probability theory to infinite sets, we are faced with the need to exercise the same kind of mathematical caution that one needs in proceeding from finite arithmetic expressions to infinite series. The "parlor game" at the beginning of Chapter 15 illustrates how easy it is to commit errors by supposing that the operations of elementary arithmetic, that are always safe on finite sets, may be carried out also on infinite sets. This includes term–by–term operations on infinite series, which may or may not yield sensible results.

In probability theory, it appears that the only safe procedure known at present is to derive our results first by strict application of the rules of probability theory on finite sets of propositions; then after the finite set result is before us, observe how it behaves as the number of propositions increases indefinitely. There are, essentially, three possibilities:

(1) It tends smoothly to a finite limit, some terms just becoming smaller and dropping out, leaving behind a simpler analytical expression.
(2) It remains bounded, but oscillates or fluctuates forever, never tending to any definite limit.
(3) It blows up, *i.e.*, becomes infinite in the limit.

In case (1) we say that the limit is "well behaved" and accept the limit as the correct solution on the infinite set. In cases (2) and (3) the limit is ill–behaved and cannot be considered a valid solution to the problem. Then we refuse to pass to the limit at all.

This is the "Look before you leap" policy: in principle, we pass to a limit only after verifying that the limit is well–behaved. Of course, in practice this does not mean that we conduct such a test anew on every problem; most situations arise repeatedly, and rules of conduct for the standard situations can be set down once and for all. But in case of doubt, we have no choice but to carry out this test.

In cases where the limit is well–behaved, it may be possible to get the correct answer by operating directly on the infinite set, but one cannot count on it. If the limit is not well–behaved, then any attempt to solve the problem directly on the infinite set would have led to nonsense, *the cause of which cannot be seen if one looks only at the limit, and not the limiting process.* The paradoxes noted in Chapter 15 illustrate some of the horrors that have resulted from carelessness in this regard.

Willy Feller on Measure Theory

In contrast to our policy, many expositions of probability theory begin at the outset to try to assign probabilities on infinite sets, both countable or uncountable. Those who use measure theory are, in effect, supposing the passage to an infinite set already accomplished before introducing probabilities. For example, Feller advocates this policy and uses it throughout his second volume (1971).

In discussing this issue (*loc. cit.*, p. 3), he notes that specialists in various applications sometimes "deny the need for measure theory because they are unacquainted with problems of other types and with situations where vague reasoning did lead to wrong results." If Feller knew of any case where such a thing has happened, this would surely have been the place to cite it – yet he does not. Therefore we remain, just as he says, unacquainted with instances where wrong results could be attributed to failure to use measure theory.

But as noted particularly in Chapter 15, there are many documentable cases where careless use of infinite sets has led to absurdities. We know of no case where our "cautious approach" policy leads to inconsistency or error; or fails to yield a result that is reasonable.

We do not use the notation of measure theory because it presupposes the passage to an infinite limit already carried out at the beginning of a derivation – in defiance of the advice of Gauss, quoted at the start of Chapter 15. But in our calculations we often pass to an infinite limit at the end of a derivation; then we are in effect using "Lebesgue measure" directly in its original meaning. We think that failure to use current measure theory notation is not "vague reasoning"; quite the opposite. It is a matter of doing things in the proper order.

Feller does acknowledge, albeit grudgingly, the validity of our position. While he considers passage to a well–defined limit from a finite set unnecessary, he concedes that it is "logically impeccable" and has "the merit of a good exercise for beginners". That is enough for us; for in this field we are all beginners. Perhaps the beginners who have the most to learn are those who decline to practice this very instructive exercise.

We note also that measure theory is not always applicable, because not all sets that arise in real problems are measurable. For example, in many applications we want to assign probabilities to functions that we know in advance are continuous. But Mark Kac (1956) notes that the class of continuous functions is not measurable; its inner measure is zero, its outer measure one.† Being a mathematician, he was willing to sacrifice some aspects of the real world in order to conform to his preconception that his sets should be measurable. So to get a measurable class of functions he enlarges it to include the everywhere discontinuous functions. But then the resulting measure is concentrated 'almost entirely' on just the class of functions that, for physical reasons, we need to exclude most strongly from our set; and so while Kac gets solutions that are satisfactory to him, they are not always the solutions to a real physical problem.

Our value judgment is just the opposite; being concerned with the real world, we are willing to sacrifice preconceptions about measurable classes in order to preserve the aspects of the real world that are important in our problem. In this case, a form of our cautious approach policy will be able to bypass measure theory in order to get the useful results we seek; for example, (1) expand the continuous functions in a finite number n of orthogonal functions, (2) assign probabilities to the expansion coefficients in a finite dimensional space R_n; (3) do the probability calculation; (4) pass to the limit $n \to \infty$ at the end. In a real problem we find that increasing n beyond a certain value makes a numerically negligible change in our conclusions (that is, if we are calculating to a finite number of decimal places, a strictly nil change). So we need never depart from finite sets.‡ Useful results, in various applications from statistical mechanics to radar detection, are found in this way.

It appears to us that most – perhaps all – of the paradoxes of infinite sets that arise in calculations are caused by the persistent tendency to pass to infinite limits too soon. In any event, whatever the cause and the cure, our position is that the paradoxes of infinite sets belong to the field of infinite set theory, and have no place in probability theory. Our self–imposed inhibition of considering only finite sets and their well–behaved limits enables us to avoid all of the useless and unnecessary paradoxing that has appeared in the recent statistical literature. From this experience, we conjecture that perhaps all correct results in probability theory are either combinatorial theorems on finite sets or well–behaved limits of them.

But on this issue, too, we are not fanatics. We recognize that the language of set and measure theory was a useful development in *terminology*, in some cases enabling one to state mathematical propositions with a generality and conciseness that is quite lacking in nineteenth century mathematics. Therefore we are happy to use that language whenever it contributes to our goal, and we could hardly get along without an occasional "almost everywhere" or "of measure zero" phrase. However, when we use a bit of measure theory, it is never in the thought that this makes the argument more rigorous; but only a recognition of the compactness of that language.

Of course, we stand ready and willing to use set and measure theory – just as we stand ready and willing to use number theory, projective geometry, group theory, topology, or any other part of mathematics – wherever this should prove helpful for the technique of finding a result or for understanding it. But we see no reason why we must state every proposition in set/measure theory terminology and notation in cases where plain English is clearer and as far as we can see, not only more efficient for our purposes but actually safer.

---

† A continuous function is defined everywhere by specifing it at each rational point, whose number is countable. Thus the class of continuous functions is very much smaller than the class of everywhere discontinuous functions.

‡ But even in the limit, the number of expansion coefficients is only countable, corresponding nicely to the property of continuous functions noted in the previous footnote.

Indeed, an insistence that all of mathematics be stated in that language all of the time can place unnecessary burdens on a theory, particularly one intended for application in the real world. It can also degenerate into an affectation, used only linguistically rather than functionally. To give every old, familiar notion a new, impressive name and symbol unknown to Gauss and Cauchy, has nothing to do with rigor. It is, more often than not, a form of gamesmanship whose real purpose is to conceal the Mickey Mouse triviality of what is being done. One would blush to state it in plain English.

Kronecker vs. Weierstrasz

At this point, a question will surely be in the reader's mind. After our emphasis on the safety of finite sets it might appear that all of analysis, which seems to do everything on uncountable sets, is suspect. Let us explain why this is not the case, and why we do place full confidence in the analysis of Cauchy and Weierstrasz.*

In the late 19'th Century both Karl Weierstrasz (1815–1897) and Leopold Kronecker (1823–1891) were at the University of Berlin,† lecturing on mathematics. A difference developed between them, which has been greatly exaggerated by later commentators, and it is only in the past few years that the real truth about their relationship has started to emerge.

Briefly, Weierstrasz was concerned with perfecting the tools of analysis – particularly power series expansions – with the specific case of elliptic functions in mind as an application. Kronecker was more concerned with the foundations of mathematics in number theory, and questioned the validity of reasoning that does not start back at the integers. On a superficial view, this might seem to deny us all the beautiful results of analysis. Even Morris Kline (1980) gives the impression that Kronecker's asceticism denies us some of the important advances in modern mathematics. But the record has been distorted.

For example, E. T. Bell (1937, p. 568) tells us, without any supporting documentation, that Kronecker on seeing Lindemann's proof that π is transcendental, asked of what use that could be, "$\cdots$ since irrational numbers do not exist?" The documentable fact is that Kronecker's own work on number theory (1901, p. 4) describes the formula of Leibniz:

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$$

as "one of the most beautiful arithmetic properties of the odd integers, namely that of determining this geometrical irrational number." Evidently, Kronecker considered irrational numbers as possessing at least enough "existence" to allow them to be precisely defined. It is true that he did not consider irrationals to be a necessary part of the foundations; indeed, how could he or anybody else think that, in view of relations like the above one, which allow irrationals to be defined entirely in terms of integers? Curiously, Weierstrasz also defined irrationals from the integers in just the same way; so where was the difference between them?

Bell also paints a picture of Weierstrasz as the great analyst, putting the final finishing touches on the work of Cauchy, and Kronecker as a mere gadfly, attacking the validity of everything he

* Indeed, the writer's first love in mathematics was not probability theory, but the use of Cauchy's complex integration to solve systems of differential equations and boundary conditions, choosing the integrand to satisfy the differential equation, and then the contour of integration to satisfy the boundary conditions. Three generations of theoretical physicists have exploited this method enthusiastically; it is great fun to teach.

† More specifically, Weierstrasz was there from 1856–1897 and Kronecker from 1861–1891. E. T. Bell (1937) gives a portrait of the young Weierstrasz and a photograph of the old Kronecker; H. M. Edwards (1989) gives photographs of the old Weierstrasz and the young Kronecker.

did without making any positive contribution. It is true that Kronecker annoyed Weierstrasz on at least one occasion, documented in Weierstrasz' correspondence; yet there was not really much conflict in their principles. To understand Kronecker's position we just need a better witness than Eric Temple Bell, and fortunately we have two of them, Henri Poincaré and Harold M. Edwards.

When Weierstrasz died in 1897, Poincaré (1899) wrote a summary of his mathematical work, in which he pointed out that: "$\cdots$ all the equations which are the object of analysis and which deal with continuous magnitudes are nothing but symbols, replacing an infinite collection of inequalities relating whole numbers." In the words of Edwards (1989), "$\cdots$ *both* Weierstrasz and Kronecker based their mathematics entirely on the whole numbers, so that all their work shared in the certitude of arithmetic." Edwards notes that several reactionary views commonly attributed to Kronecker are hearsay for which no support can be found in Kronecker's own words.

The difference between Kronecker and Weierstrasz was aesthetic rather than substantive; Kronecker wants to keep first principles (the origin in the integers) constantly in view, while Weierstrasz, having made a new construction, is willing to forget the steps by which it was made, and use it as an element in its own right for further construction. Put in modern computer terminology, Weierstrasz did not deny Kronecker's "machine language" basis of all mathematics, but wanted to develop analysis in a higher level language. Edwards points out that Kronecker's principles, "$\cdots$ in his mind and in fact, were no different from the principles of his predecessors, from Archimedes to Gauss."

Thanks largely to the historical research of Edwards, the truth is emerging and Kronecker is being vindicated and rehabilitated. Perhaps Kronecker was overzealous, and perhaps he misunderstood the position of Weierstrasz; but events since then suggest that he was not zealous enough in his own cause. His failure to respond to Cantor seems unfortunate, but easy to understand.

To Kronecker, Cantor's ideas were so *outré* that they had nothing to do with mathematics, and there was no reason for a mathematician to take any note of them. If the editors of the mathematical journals made the mistake of publishing such stuff, that was their problem, not his. But the messages that Kronecker did communicate contained some very important truth; in particular he complained that much of set theory was fantasy because it was not algorithmic (that is, it contained no rule by which one could decide, in a finite number of operations, whether a given element did or did not belong to a given set). Today, with our computer mentalities, this seems such an obvious platitude that it is hard to imagine anyone ignoring it, much less denying it; yet that is just what happened. We think that, had mathematicians paid more attention to this warning of Kronecker, mathematics might be in a more healthy state today.

What is a Legitimate Mathematical Function?

Much of the difference between current pure and applied mathematics lies in their different conceptions of the notion of a "function". Historically, one started with the well–behaved analytic functions like $f(x) = x^2$ or $f(x) = \sin x$. Then these were generalized, but in two different ways. In pure mathematics, the idea was generalized in such a way that set theory notions remained valid; first to piecewise continuous functions, then to quite arbitrary rules by which, given a number x, one can define another number f. Then, perceiving that a function or its argument need not be limited to real or complex numbers, this was generalized to an arbitrary mapping of one set X onto another set F, the elements of which could be almost anything.

In applied mathematics, the notion of a function was generalized in a very different way; so that the useful *analytical operations* that we perform on functions remain valid. Perhaps the most important hint was provided by the operation of the fourier transform. This is still a mapping, but at the higher level of mapping one function $f(x)$ onto another $F(k)$. This mapping was defined by the integrals

$$F(k) = \int e^{ikx} f(x)\, dx\,, \qquad f(x) = \frac{1}{2\pi} \int e^{-ikx} F(k)\, dk\,. \tag{B–6}$$

If we indicate this fourier transform pair symbolically as

$$\Big[f(x) \leftrightarrow F(k) \Big] \tag{B–7}$$

we find the interesting properties that under translation and convolution

$$\Big[f(x-a) \leftrightarrow e^{ika}\, F(k) \Big] \tag{B–8}$$

$$\Big[\int f(x-y)\, g(y)\, dy \leftrightarrow F(k)\, G(k) \Big] \tag{B–9}$$

and under differentiation,

$$\Big[f'(x) \leftrightarrow ikF(k) \Big]\,, \qquad \Big[-ix\, f(x) \leftrightarrow F'(k) \Big] \tag{B–10}$$

if the integrals converge. In other words, analytical operations on one function correspond to algebraic operations on the other.

In practice these are very useful properties. Thus to solve a linear differential equation, or difference equation, or integral equation of convolution form $[\int K(x-y)\, f(y) dy = \lambda\, g(x)]$ or, indeed a linear equation which contains all three of these operations, one may take its fourier transform, which converts it into an algebraic equation. This may be solved directly. Then taking the inverse fourier transform yields the solution of the original equation. Thus the fourier transform mapping reduces the solution of linear analytical equations to that of ordinary algebraic equations. The theoretical physicist Arnold Sommerfeld in Munich became a great artist in the technique of evaluating these solutions by fancy contour integrals, and some of the greatest of the next generation learned this from him. Today, physicists and engineers could hardly survive without it.

But this procedure seemed to apply only to a limited class of functions. In the Dirichlet form of fourier theory, one shows that if $f(x)$ is absolutely integrable, then the integral (B–6) surely converges to a well–behaved continuous function $F(k)$ on the real axis, and all is well. If $f(x)$ vanishes for negative x, then $F(k)$ is analytic and bounded in one half of the complex plane, and all is even better. But if $f(x)$ is absolutely integrable, then $f'(x)$ or $f''(x)$ may not be; and there is some doubt whether the useful properties are still valid. In the early work on fourier transforms, such as Titchmarsh (1937), virtually all one's attention is concentrated on the theory of convergence of the integrals, and any function for which the integral did not converge was held not to possess a fourier transform. This placed an intolerable restriction on the range of useful applications of fourier theory.

But then a more sophisticated view emerged in theoretical physics. One realized that the usefulness of the fourier transform lies, not in convergence of any integral, but in the above properties (B–8) – (B–10) of the mapping. Therefore if by any means we can define a mapping such that these properties are preserved, then the customary use of fourier transforms to solve linear integrodifferential equations will be perfectly rigorous and *it does not make the slightest difference* whether the integrals (B–6) or the analogous fourier series do or do not converge. A divergent fourier series is still a unique ordered sequence of numbers, conveying all the needed information (tbat is, it is uniquely determined by, and uniquely determines, its Fourier transform). It was only an historical accident that this mapping was first discovered through their series and integral representations, which exist only in special cases.

Although its beginnings can be traced back to Duhamel and Green in the 19'th Century, this movement is commonly held to start with P. A. M. Dirac, who in the 1920's invented the notation of the delta–function $\delta(x-y)$ and showed how to use it to good advantage in theoretical physics. It is the "fourier transform of a constant" in the sense that as $F(k) \rightarrow 1$, we have $f(x) \rightarrow \delta(x)$. Mathematicians thinking in terms of the set theory definition of a 'function' were horrified and held this to be nonrigorous on the grounds that delta–functions do not 'exist'. But that was only because of their inappropriate definitions of the term 'function'. A delta–function is not a mapping of any set onto any other. Laurent Schwartz (1950) tried to make the notion of a delta-function rigorous, but from our point of view awkwardly, because he persisted in defining the term 'function' in a way inappropriate to the subject.

Perceiving this, G. Temple (1955) and M. J. Lighthill (1957) showed how to remove the awkwardness simply by adopting a definition of functions as meaning "good" functions and limits of sequences of good functions (thus in our system, a discontinuous function is *defined* as the limit of a sequence of continuous functions). For this, there is almost no need to mention such things as open and closed sets. Lighthill saw that this definition of 'function' is the one appropriate to fourier theory. It is now clear that it is also the one appropriate to probability theory and to all of analysis; with it our theorems become simpler and more general, without a long list of exceptions and special cases. For example, any fourier series may now be differentiated term by term any number of times and the result, whether convergent or not, identifies (by 1:1 correspondence) a unique function in our sense of the word. Physicists had seen this intuitively and used it correctly long before the work of Temple and Lighthill.

Lighthill produced a very thin book (1957) on the new form of fourier analysis, which included a table of fourier transforms in which every entry is a function which was held formerly not to possess a fourier transform. Yet that table is a gold mine for the useful solution of linear integro–differential equations. In a famous review of Lighthill's book, the theoretical physicist Freeman J. Dyson (1958), a former student of the Cambridge mathematician G. H. Hardy, said that Lighthill's book " $\cdots$ lays Hardy's work in ruins, and Hardy would have enjoyed it more than anybody." Throughout the present work we take Lighthill's approach, as summarized briefly in Appendix F, for granted and assume that the reader is familiar with it.

Nondifferentiable Functions

The issue of nondifferentiable functions arises from time to time in probability theory. In particular, when one solves a functional equation like those studied in Chapter 2, to assume differentiability is to have a horde of mathematical nit–pickers descend upon one, with claims that we are excluding a large class of potentially important solutions. However, we noted that this is not the case; Aczel demonstrated that Cox's functional equations can all be solved without assuming differentiability (at the cost of much longer derivations) and with just the same solutions that we found above. Then is it possible that this genre of mathematical nit–picking is, like the rejection of delta–functions, only an artifact of inappropriate definition of terms?

Let us take a closer look at the notion of nondifferentiable functions in general. This was not well received at first by pure mathematicians. Charles Hermite wrote to Stieltjès: "I turn away in horror from this awful plague of functions which have no derivatives." The one generally blamed for this plague was Lebesgue, although Weierstrasz had noted them before him. The Weierstrasz nondifferentiable function is

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(m^n \pi x) \tag{B–11}$$

where $0 < a < 1$) and m is a positive odd integer. The series is uniformly convergent for all x (since it must converge at least as well as does $\sum a^n$), so it defines a continuous function. But if $ma > 1$, term–by–term differentiation yields a divergent series. The proof that the derivative

$$f'(x) \equiv \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{B–12}$$

then does not exist is rather tedious (G. H. Hardy, Proc. London Math. Soc.(2), **9**, pp. 126–144). Titchmarsh (1939, pp. 350–353) gives only a shorter proof valid when $ma > 1 + 3\pi/2$. Weierstrasz' function is, in fact, the limit of a sequence of good functions (the partial sums S_k of the first k terms), and so we might admit it into our system; but it is not a very well–behaved limit.

Fake Nondifferentiable Functions The case most often cited as an example of a nondifferentiable function is derived from the sequence $f_n(x)$, a string of tiny right triangles whose hypotenuses lie on the real axis and have length $1/n$. For any n, the slope of $f_n(x)$ is ± 1 almost everywhere. Then what happens as $n \to \infty$? The limit $f_\infty(x)$ is often cited carelessly as a nondifferentiable function. Now it is clear that the limit of the derivative, $f'_n(x)$ does not exist; but it is the derivative of the limit that we should examine; $f_\infty(x) \equiv 0$ and this is certainly differentiable; $f'_\infty(x) = 0$ everywhere. Many other such sequences $f_n(x)$ with discontinuous slope on a finer and finer scale may be defined. The error of calling the limit $f_\infty(x)$ nondifferentiable on the grounds that the limit of the derivative does not exist, is rampant in the mathematical literature. In these cases, the limit of a sequence of bad functions is actually – contrary to what is often stated – a good function and we have no reason to exclude it from our system.

Lebesgue defended himself against his critics thus: "If one wished always to limit himself to the consideration of well–behaved functions, it would be necessary to renounce the solution of many problems which were proposed long ago and in simple terms." The present writer is unable to cite any specific problem which was thus solved; but we can borrow Lebesgue's argument to defend our own position.

To reject limits of sequences of good functions is to renounce the solution of many current real problems. Those limits can and do serve many useful purposes, which much current mathematical education and practice still tries to stamp out. Indeed, the refusal to admit delta–functions as legitimate mathematical objects has led mathematicians into error. For example, H. Cramér (1946, Chap. 32) gives an inequality, which we derived in Chapter 17, placing a lower limit to the variance of the sampling distribution for a parameter estimator θ^*:

$$var(\theta^*) \ge \frac{(1 + db/d\theta)^2}{n \int (\partial \log f/\partial\theta)^2 \, f(x|\theta)\, dx} \tag{B–13}$$

where we have made n observations from a sampling distribution $f(x|\theta)$, and $b(\theta^*) \equiv E(\theta^* - \theta)$ is the bias of the estimator.

Then Cramér notes that if $f(x|\theta)$ has discontinuities, then "the conditions for the regular case are usually not satisfied. In such cases it is often possible to find unbiased estimates of 'abnormally high' precision, *i.e.*, such that the variance is smaller than the lower limit [(B–13)] for regular estimates."

How could he have reached such a remarkable conclusion, since (B–13) is only the Schwartz inequality, which does not seem to admit of exceptions? We find that he has used the set–theory definition of a function, and concluded that the derivative $\partial \log f/\partial\theta$ does not exist at points of discontinuity. So he takes the integral in (B–13) only over those regions where $f(x|\theta)$ is continuous.

But the definition of a discontinuous function which is appropriate in analysis is our limit of a sequence of continuous functions. As we approach that limit, the derivative develops a higher and sharper spike. However close we are to that limit, the spike is part of the correct derivative of the function, and its contribution must be included in the exact integral. Thus the derivative of a discontinuous function necessarily contains a delta–function at points of discontinuity, *whose contribution is always present in the differentiated fourier series, and must be included in order* *to get the correct physical solution.* Had Cramér included this term, there would have been no anomalous case, and no seeming violation of the Schwartz inequality. In this case, a disbelief in delta–functions has led a mathematician into an elementary error in calculation.

In a similar way, the solution of an integral equation with finite limits, of the form

$$\int_a^b K(x,y) f(y) dy = \lambda g(x) \tag{B–14}$$

generally involves delta–functions like $\delta(y-a)$ or $\delta'(y-b)$ at the end–points, and so those who do not believe in delta–functions consider such integral equations as not having solutions. But in real physical problems, exactly such integral equations occur repeatedly, and again the delta–functions must be included in order to get the correct physical solution. Some examples are given by D. Middleton (1960), and we hope to give many more in the planned second volume of this work. It is astonishing that so few people have yet perceived this, but we think it only illustrates what we have observed independently; those who think of fundamentals in terms of set theory almost never get around to useful analytical calculations.

Yet nondifferentiable functions are manufactured as limits of sequences of such things as rows of tinier and tinier triangles, and this is accepted without complaint, as generating an important new class of functions. Those who do this while looking askance at delta functions, are in the position of admitting limits of sequences of bad functions as legitimate mathematical objects, while refusing to admit limits of sequences of good functions! This seems to us a sick policy, for delta–functions serve many essential purposes in real, substantive calculations, but we are unable to conceive of any useful purpose that could be served by a nondifferentiable function. Their only use is to provide trouble makers with artificially contrived counter–examples to almost any sensible and useful mathematical statement one could make. Henri Poincaré (1909) noted this in his characteristically terse way: "*In the old days when people invented a new function they had some useful purpose in mind: now they invent them deliberately just to invalidate our ancestors' reasoning, and that is all they are ever going to get out of them.*"

Indeed, this fad of artificially contrived mathematical pathology seems nearly to have run its course, and for just the reason that Poincaré foresaw; nothing useful can be done with it. While we still see exhortations not to assume differentiability of an unknown function, on grounds of loss of generality, it is difficult to find even one specific example of a nondifferentiable function appearing – much less actually being used for anything – in the recent literature. To the best of our knowledge, they play no role in the discussions of any mathematical question or useful application and one must go back to old works like Titchmarsh (1939) to see them at all.

Note, therefore, that we stamp out this plague too, simply by our defining the term "function" in the way appropriate to our subject. The definition of a mathematical concept that is 'appropriate' to some field is the one that allows its theorems to have the greatest range of validity and useful applications without the need for a long list of exceptions, special cases, and other anomalies. In our work the term 'function' includes good functions as defined in Appendix F, and well–behaved limits of sequences of good functions; but not nondifferentiable functions. We do not deny the existence of other definitions which do include nondifferentiable functions, any more than we deny the existence of fluorescent purple hair dye in England; in both cases, we simply have no use for it.†

Finally, we offer a few more general comments on mathematical style.

† On a different topic, we follow the same policy by defining the term "moving average" for a finite time series in such a way that our theorems are all exact, without any need for messy "end effect" corrections. Of course, it then develops that this is the definition most directly useful in applications (Chapter 23).

What Am I Supposed to Publish?

L. J. Savage (1962) asked this question to express his bemusement at the fact that, no matter what topic he chose to discuss and no matter what style of writing he chose to adopt, he was sure to criticized for not making a different choice. In this he was not alone. We would like to plead for a little more tolerance of our individual differences.

If anyone wants to concentrate his attention on infinite sets, measure theory, and mathematical pathology in general, he has every right to do so. And he need not justify this by pointing to useful applications or apologize for the lack of them; as was noted long ago, abstract mathematics is worth knowing for its own sake.

But others in turn have equal rights. If we choose to concentrate on those aspects of mathematics which *are* useful in real problems and which enable us to carry out important substantive calculations correctly – but which the mathematical pathologists never get around to – we feel free to do so without apology.

Ultimately, the mathematical level and depth of this work were chosen with the aim of making it possible for all readers to get what they want from it. Since those who approach a work with the sole purpose of finding fault with its style of presentation will always be able to do so no matter how it is presented, our aim was to ensure that those who approach it with sincere desire to understand its *content* will also be able to do so. Thus we try to give cogent reasons why the ideas we advocate are "obvious," while those we deplore are not, when this can be done briefly enough not to interrupt the line of argument. This inevitably leaves some lacunae, in part filled in by the Comments at the end of most Chapters.

In this connection, the question of what is and is not "obvious" is a matter of gamesmanship that is played in two opposite directions. On the one hand, the standard way of introducing notions that do not stand up to critical examination – or to deprecate those that stand up too well to be safely opposed – is to call them "obvious." On the other hand, to express grave doubts about simple matters that *are* obvious, is the equally standard technique for imputing to one's self deep critical faculties not possessed by others. We try to steer a middle course between these, but like Savage do so in the knowledge that whatever our choice, it will receive opposite criticisms from the two types of gamesman.

But we avoid one common error; nothing could be more pathetically mistaken than the prefatory claim of one author in this field that mathematical rigor "guarantees the correctness of the results." On the contrary, much experience teaches us that the more one concentrates on the appearance of rigor, the less attention he pays to the validity of the premises in the real world, and the more likely he is to reach final conclusions that are absurdly wrong in the real world.

Mathematical Courtesy

A few years ago the writer attended a Seminar talk by a young mathematician who had just received his Ph.D. degree and, we understood, had a marvelous new limit theorem of probability theory. He started to define the sets he proposed to use, but three blackboards were not enough for them and he never got through the list. At the end of the hour, having to give up the room, we walked out in puzzlement, not knowing even the statement of his theorem.

A "Nineteenth Century (C19) Mathematician" like Poincaré would have been into the meat of the calculation within a few minutes and would have completed the proof and pointed out its consequences in time for discussion.

The young man is not to be blamed; he was only doing what he had been taught a "Twentieth Century (C20) Mathematician" must do. Although he has perhaps now learned to plan his talks a little better, he is surely still wasting much of his own time and that of others in reciting all the preliminary incantations that are demanded in C20 mathematics before one is allowed to proceed to the actual problem. He is a victim of what we consider to be, not higher standards of rigor, but studied mathematical discourtesy.

Nowadays, if you introduce a variable x without repeating the incantation that it is in some set or "space" X, you are accused of dealing with an undefined problem. If you differentiate a function $f(x)$ without first having stated that it is differentiable, you are accused of lack of rigor. If you note that your function $f(x)$ has some special property natural to the application, you are accused of lack of generality. In other words, every statement you make will receive the discourteous interpretation.

Obviously, mathematical results cannot be communicated without some decent standards of precision in our statements. But a fanatical insistence on one particular form of precision and generality can be carried so far that it defeats its own purpose. C20 mathematics often degenerates into an idle adversary game instead of a communication process.

The fanatic is not trying to get your substantive message at all, but only trying to find fault with your style of presentation. He will strive his utmost to read nonsense into what you are saying, if he can possibly find any way of doing so. In self–defense, writers are obliged to concentrate their attention on every nit–picking detail of how things are said rather on what is said. The length grows, the content shrinks.

Mathematical communication would be much more efficient and pleasant if we adopted a different attitude. For one who makes the courteous interpretation of what others write, the fact that x is introduced as a variable *already implies* that there is some set X of possible values. Why should it be necessary to repeat that incantation every time a variable is introduced, thus using up two symbols where one would do? [Indeed, the range of values is usually indicated more clearly at the point where it matters; by adding conditions like $(0 < x < 1)$ after an equation].

For a courteous reader, the fact that a writer differentiates $f(x)$ twice already implies that he considers it twice differentiable; why should he be required to say everything twice? If he proves proposition A in enough generality to cover his application, why should he be obliged to use additional space for irrelevancies about the most general possible conditions under which A would be true?

C19 mathematicians were not being nonrigorous by their style; they merely, as a matter of course, extended simple civilized courtesy to others, and expected to receive it in return. This will lead one to try to read sense into what others write, if it can possibly be done in view of the whole context; not to pervert our reading of every mathematical work into a witch–hunt for deviations from the Official Style.

Therefore, sympathizing with the young man's plight but not intending to be enslaved like him, we issue the following:

EMANCIPATION PROCLAMATION

Every variable x that we introduce is understood to have some set X of possible values. Every function $f(x)$ that we introduce is understood to be sufficiently well behaved so that what we do with it makes sense. We undertake to make every proof general enough to cover the application we make of it. It is an assigned homework problem for the reader who is interested in the question, to find the most general conditions under which the result would hold.

We could convert many C19 mathematical works to C20 standards by making a rubber stamp containing this Proclamation, with perhaps another sentence using the terms "Sigma–algebra, Borel field, Radon–Nikodym derivative", and stamping it on the first page.

Modern writers could shorten their works substantially, with improved readability and no decrease in content, by including such a Proclamation in the Copyright message, and writing thereafter in C19 style. Perhaps some publishers, seeing these words, may demand that they do this for economic reasons; it would be a service to science.

In this Appendix we have presented many short quotations without the references. Supporting documentation and many further interesting details may be found in Bell (1937), Félix (1960), Kline (1980), and Rowe & McCleary (1989).