A FEW SEEDLINGS OF RESEARCH

J. M. HAMMERSLEY
UNIVERSITY OF OXFORD

1. Sowing

Graduate students sometimes ask, or fail to ask: "How does one do research in mathematical statistics?" It is a reasonable question because the fruits of research, lectures and published papers bear little witness to the ways and means of their germination and ripening. How did that author ever come to state this theorem so aptly, to snatch this neat proof from the thin air: surely it never sprang fully armed like Pallas Athene from the brow of Zeus? No, indeed not, rest reassured. But still the question is a hard one. The answer depends much upon the field of statistics, and even more upon the tastes and skills and prejudices of the researcher. The only means of appraisal is case study, and no author has the data for any case study but his own. Without more ado or apology, I shall speak of some of my own work and how it came about; and it will have to be a personal story, with all natural drawbacks, and of course to get a proper view of things at large, you will, nay must, look elsewhere for other accounts by other men of their own work.

The Committee on Support of Research in the Mathematical Sciences, in the introduction to a volume of essays [21], wrote:

"Our task was to assess the present status and the projected future needs, especially fiscal needs, of the mathematical sciences. . . . We realize that even scientific readers of our report, let alone nonscientists, may feel that they are not adequately informed about what mathematical research, especially modern mathematical research, consists of. Similarly, even professional mathematicians . . . may be unaware of the applications. . . . To provide additional background of factual information concerning the mathematical sciences, we are supplementing our report with the present collection of essays. . . . We believe that the mathematical community has no obligations more important than those concerned with education. . . ."

The essays in this collection provide an excellent account of the substance and applications of mathematical research: it deals, as the Committee intended, with the what. It does not try to deal with the how. Yet anyone agreeing with the last sentence in the above quotation, as I heartily do, must feel that the how also deserves educational coverage. So I have conceived the present paper as if it were *one* chapter for a companion volume, yet to be compiled, on the how of mathematical research.

The reader whom I have in mind, mainly though not exclusively, is a graduate student just about to embark on research in mathematical statistics. He will have had a thorough undergraduate education in mathematics, so he will have a good technical grasp of the subject; but probably he will also be limited by some of the unavoidable drawbacks inherent in any undergraduate course. In particular, he will suffer from a surfeit of knowledge and from overexposure to the *fait accompli* and to rigor. In the secondary school he will have done some nonrigorous mathematics (at least I hope so) if he has not squelched too deeply into the "New Math" mire; but he will not yet know how to argue nonrigorously at a more advanced level. So I have deliberately written a good deal of this paper in a nonrigorous style; for nonrigorous thought is an essential first stage in almost any piece of research. The rigor comes at the end of the process: undergraduate training serves the end, not the start.

Most of this paper deals with one particular problem: the problem itself, namely the what, is less important than its how. It would be impossibly tedious to give the how in full detail; and there is no need for me to say much about the false avenues explored or the mistakes made. Everybody commits these, and no one finds them hard to imagine. What I have done instead, and what I hope will suffice, is to leave the investigation of the problem in an unfinished state. This may show in snapshot form what research looks like while still in progress, not as it usually appears in the literature as a packaged end-product. The material in the later parts of the paper is at a more rudimentary stage of development than in the earlier parts, where I have ventured to formulate some of the results as theorems. This makes a virtue of necessity; for I have had to write for a publication deadline. The whole investigation has occupied eleven weeks—about three weeks to do the mathematics, interspersed throughout eight weeks of writing it up. It will be no bad thing for realism if this haste shows in the text.

There are two other things. First, I have emphasized, especially at the end of the paper, conjectures and unsolved problems. These, into which the graduate student may feel inclined to stick his teeth, serve the same purpose as exercises in a textbook. Second, I have posed a problem (on the interfusion of aluminium and copper) more or less in the raw state in which it comes from the scientist to the mathematician. This can serve as an exercise in constructing a suitable mathematical model (the one sketched in the text may *not* be suitable).

Some people would say that recent years have been a golden age for mathematical research, with funds and graduate students on an unprecedented scale. I thought so too once; but, having seen the age unfold, my views have changed. The plenitude has enlarged the quantity of mediocre research without enhancing quality. I regret the way a doctorate has become virtually a union ticket for university employment. Admittedly, my own experience colors these views: I was lucky enough never to have to write a dissertation. Having started before the time when doctorates were fashionable, I was able to jump in at the deep end straight away; and, instead of being restrained by the single theme of a dissertation, I could pick and choose from a wide range of scientific enquiries, of a richness which today's graduate students rarely meet or even visualize, and in the three years it would have taken me to write one doctoral thesis I published a dozen papers on all sorts of different topics.

Who sows the seed? Whence the subject matter? By inclination I am not a theorist, but a problem solver. When I first acquired my bachelor's degree in mathematics some twenty years ago, I had the good fortune to be appointed as a junior assistant in a small department at Oxford whose purpose and name was the Design and Analysis of Scientific Experiment. This meant it had to do the sums and headscratching and minister to the general statistical needs of biologists, foresters, and farmers in the university. There could have been no better place for source material. Scientists of all breeds would bring in a cornucopia of subjects, susceptible of mathematical treatment. This is not to claim that I regularly answered their questions: far from it, a mathematician in these circumstances will more likely convert the questions to his own purposes. For example, quite early on I managed to conjure a paper on nonharmonic Fourier series [10] out of the apparently unpromising raw material of carrot roots soaked in acetic acid, much to the consternation of the botanist concerned. (This may be the only occasion when the august pages of *Acta Mathematica* have talked of carrots and vinegar.) Later I moved towards theoretical physics, which is a chastening experience because theoretical physicists tend to be much better at mathematics than professional mathematicians themselves. (Sykes and Essam's work, to be mentioned in Section 20, is a good example of the mathematical virtuosity of physicists.) After some years, I got known for having solved a few of the problems put to me. This, let me explain with all due lack of modesty, is an unusual reputation for any mathematician to earn. Accredited mathematical statisticians never—what never? well, hardly ever— solve any problems at all: for, no sooner do they really acquire that knack, than they become known as theoretical physicists, or molecular biologists, or engineers, or the like; but I cannot claim to have been so successful as that. Nevertheless, over the years I have accumulated a stock of unsolved or partly solved mathematical problems.

So I could rummage in this stock for something to talk about when invited to speak at the Sixth Berkeley Symposium; and presently I came across the problem of what will (or might) happen if you clamp a flat sheet of copper closely against an aluminum one. As a simplified mathematical model, we can think of the copper sheet as occupying one half of space, and the aluminum the other half, with a plane boundary in between. Each metal consists of atoms positioned on a lattice: say, for simplicity, the lattice is the set of points with integer coordinates (x, y, z). Initially the copper atoms occupy the lattice points with $x \geqq 0$ while the aluminum occupies the lattice points with $x < 0$. However, a small percentage of the lattice points have no atom: these are called vacancies. The situation is a three-dimensional analogue of the child's toy with fifteen small numbered squares which can slide in a square box large enough for sixteen. Any of the six atoms next to a vacancy may by chance fall into it; and if some of these six atoms are copper and the rest aluminum, one kind may enjoy a preferential probability of jumping into the vacancy. Naturally, when an atom jumps into a vacancy, it leaves a new vacancy whence it came: or in other words, the vacancy jumps to a neighboring lattice point in exchange for an atom. Thus the vacancies describe random walks over the lattice. Yet this motion of the vacancies is only of intermediate concern: the ultimate interest is the motion they impart to the atoms. How will a marked atom move with the passage of time? Or better, what are the equations of the general atomic commotion caused by the randomly walking vacancies, and how fast will the copper plate diffuse into the aluminum plate and vice versa? How will this depend upon the percentage concentration of vacancies and any preferential probabilities postulated?

And now, dear graduate student, I pass this problem on to you as a possible topic for research, because with only two weeks to go before the Berkeley Symposium my own analysis of it was suddenly interrupted by that indefatigable beaver, Professor Gian-Carlo Rota of M.I.T. Might he jog my memory (he wrote firmly) of the collected works of Ulam under his editorship, and of its imminent dispatch to the printers, and of my promise to write some comments on certain papers he had sent me? And so I turned to these papers by Ulam, and at once a problem in one of them [25] intrigued and fascinated me.

That is how the seeds of research may be sown. Chance sows, and curiosity nurtures.

2. Ulam's problem

Ulam asks what is the distribution of the length (that is, number of terms) of a longest monotone subsequence of (not necessarily consecutive) terms in a random permutation of the first $n^2 + 1$ natural numbers. For example, if $n = 2$ and $n^2 + 1 = 5$, the permutation 51423 contains no monotone subsequence of length 4, but at least one (actually three) of length 3 (.1.23, 5.42., and 5.4.3), so any longest monotone subsequence of this particular permutation has length $3 = n + 1$. There are 120 possible permutations of 12345; and the longest length is 3 for 86 of them, 4 for 32 of them, and 5 for the remaining 2.

Ulam remarks that the longest length must always be at least $n + 1$ for any n, by virtue of what he calls "a well-known theorem." He then refers to a Monte Carlo study of the cases $n = 4$ to $n = 10$ by E. Neighbor: he cites no reference, and indeed Neighbor's work may be unpublished. For three of these values of n, he quotes the average longest lengths in the Monte Carlo samples, namely,

$$\begin{aligned} 8.46 &= 1.69n \quad \text{for} \quad n = 5 \\ 14.02 &= 1.75n \quad \text{for} \quad n = 8 \\ 17.85 &= 1.78n \quad \text{for} \quad n = 10. \end{aligned} \tag{2.1}$$

All the averages, he says, were about $1.7n$, while the distribution "turned out to have a Gaussian form starting at the guaranteed minimum [that is, $n + 1$], having its maximum at the average, and becoming vanishingly small at about 2.2 times the minimum." A careless, or perhaps a fancifully numerological reading of this quotation might suggest that the distribution was spread between $n\sqrt{1}$ and $n\sqrt{5}$ with mean $n\sqrt{3}$, thus giving a standard deviation also proportional to n. We shall discover that all these suggestions are wrong.

Now, a highly finished paper, with all its theorems carefully proved, all avenues explored, and all loose ends carefully snipped, may arouse one's admiration; but its very perfection drains it of vitality, and there is little one can do with it except file it. Papers are more entertaining if they are still rich in conjectures, with results unproved or even wrong: Ulam's paper is like this; and I shall do what I can to write here in a similar vein, incidentally saving space by avoiding rigor. As a high exemplar, we may recall that the most enigmatically stimulating communication in the whole history of mathematics was written under extreme restrictions on space—margin too small.

3. How well known is a well-known theorem!

Ulam's "well-known theorem" runs as follows:

THEOREM 1. (Ulam? or A.N. Other??) *Any real sequence of* $n^2 + 1$ *terms contains a monotone subsequence of* $n + 1$ *terms.*

Enterprising readers will doubtless wish to prove this theorem for themselves: so I defer the proof while they get to work. I regret to say this theorem was not well-known to me, indeed not known at all; and it took me a couple of days to invent my own proof, based on negative induction (if the theorem is false for n, it is also false for $n - 1$). Having found a proof, I cast around for colleagues who might know the theorem and be able to cite a reference (Ulam gives neither proof nor reference, thereby arousing his reader's curiosity). It was a long time before I met anyone able to answer: eventually, however, I turned up trumps in Professor Lincoln Moses. He believed there was a proof, based on the pigeonhole principle, in a collection of essays [21] prepared for the Committee on Support of Research in the Mathematical Sciences (COSRIMS); and he thought that Ulam himself had written the essay in question. Together Moses and I went to the Berkeley Library to verify this reference: but fortunately we found it was out on loan and so temporarily unavailable. I say "fortunately" because I now had the chance of reconstructing on my own a second proof, using the pigeonhole principle: and the reader has too.

A few days later I delivered the lecture on this paper at a session of the Sixth Berkeley Symposium; and I took the opportunity of doing a little operational research on the familiarity of well-known theorems. I asked the audience how many of them knew Ulam's "well-known theorem." Out of an audience of 59 accomplished mathematicians, just 3 knew it. Adding 1 to 59 to include my own ignorance, we conclude that 95 percent of mathematicians will be ignorant of a well-known theorem. But one theorem is a small sample of theorems.

Professor Milton Sobel was one of the three wise men in the audience at the Berkeley Symposium; and he told me that the theorem plus a proof—indeed a a third way of proving it—had appeared in Martin Gardner's column [7] in

Scientific American. It is astonishing that the theorem should be so little known after appearance in such justly famous and widely read columns.

Lastly, the COSRIMS essays [21] came back to the library, and I was able to consult them. And it came to pass as Moses had prophesied, that Ulam had indeed contributed an essay to this collection; but his essay made no mention of the theorem. Yet the collection contained another essay, which did mention it; and—curious coincidence?—the author of this essay was Gian-Carlo Rota, who, as we have already seen, is guilty of this whole diversion away from copper and aluminum.

The purpose of this chit-chat on the background to Theorem 1 has been to interpose a long enough piece of prose to discourage the reader's eye from straying from the statement of the theorem to its proof. But it also serves to introduce a more serious maxim, which every properly ambitious graduate student should know. For Rota's use of the pigeonhole principle, when I eventually read it, did not seem to me quite as satisfactory as the one I had reconstructed for myself: his argument is somewhat longer and proceeds by *reductio ad absurdum*, whereas my more direct line is a constructive proof. Moral: never read the literature before you absolutely have to (and not always, even then), for thus it will not cloud your imagination and sometimes you may be able to do better on your own.

Added in in proof. Theorems 1 and 2 are both due to Erdös and Szekeres *Compositio Math.*, Vol. 2 (1935), pp. 463–470 (in particular, page 468).

Theorem 1 is the particular case $a = d = n$ of the slightly more general Theorem 2 (which could be new, though I doubt it).

THEOREM 2. *Any real sequence of at least* $ad + 1$ *terms contains either an ascending subsequence of* $a + 1$ *terms or a descending subsequence of* $d + 1$ *terms.*

Here and later I interpret "ascending" to mean "nondecreasing," and "descending" to mean "strictly decreasing." (Theorem 2 is also true for "ascending" = "strictly increasing," and "descending" = "nonincreasing," the changes in the proof being trivial.)

Suppose we have a set of pigeonholes $P_1, P_2, \cdots$ and a given real sequence $X_N = \{x_1, x_2, \cdots, x_N\}$ with $N \geqq ad + 1$. Place the terms of X_N successively in the pigeonholes: first put x_1 in P_1; then generally, with $x_1, x_2, \cdots, x_{i-1}$ already placed, place x_i in the pigeonhole P_j with the least value of j such that P_j already contains no term larger than x_i. Thus, at any stage of the procedure the contents of each pigeonhole comprises an ascending subsequence of X_N. Moreover (since j is least), when x_i goes into P_j, then P_{j-1} must already contain an earlier and greater term (that is, some x_h such that $h < i$ and $x_h > x_i$); and similarly P_{j-2} must contain an earlier and greater term than this x_h, and so on as far as P_1. Hence, there is a descending sequence with one term in each occupied pigeonhole (and it is explicitly recoverable by working backwards from any term in the last occupied pigeonhole). The required subsequence will have been constructed as soon as either some pigeonhole contains $a + 1$ terms or $d + 1$ pigeonholes become occupied. This event will occur because of the pigeonhole principle and because $N \geqq ad + 1$.

4. Notation

The following notation will be useful. If r is any real number, define $\chi(r)$ to be the least integer such that $\chi(r) \geqq r$. We write

$$X = \{x_1, x_2, \cdots\} \tag{4.1}$$

for an infinite real sequence; and

$$X_N = \{x_1, x_2, \cdots, x_N\} \tag{4.2}$$

for the first N terms of X. We write

$$\ell_N = \ell(X_N) \tag{4.3}$$

for the length of a longest ascending subsequence of X_N, and

$$\ell'_N = \ell'(X_N) \tag{4.4}$$

for the length of a longest descending subsequence of X_N. Thus

$$\ell^*_N = \ell^*(X_N) = \max(\ell_N, \ell'_N) \tag{4.5}$$

denotes the length of a longest monotone subsequence of X_N.

With this notation we can now state Theorem 3.

THEOREM 3. *For any real sequence* X,

$$\ell^*(X_N) \geqq \chi(\sqrt{N}), \tag{4.6}$$

and this inequality is best possible.

Since $\ell^*(X_N)$ is a nondecreasing function of N, the inequality (4.6) follows at once from the particular case $N = n^2 + 1$ covered by Theorem 1. To see that (4.6) is best possible, we arrange the positive integers into successive parts of one, two, three, $\cdots$ terms; and we reverse each part to obtain (1) (3 2) (6 5 4) (10 9 8 7) $\cdots$. Then we interlace this sequence with its negative to yield the particular sequence $X = \{x_1, x_2, \cdots\}$ exhibited in (4.7).

(4.7)

N:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	$\cdots$
x_N:	1	−1	3	2	−3	−2	6	5	4	−6	−5	−4	10	9	8	7	−10	$\cdots$
ℓ_N:	1	1	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	$\cdots$
ℓ'_N:	1	2	2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	$\cdots$.

We can now verify for this particular sequence X that

$$\ell_N = \chi[(N + \tfrac{1}{4})^{1/2} - \tfrac{1}{2})] \leqq \ell'_N = \ell^*_N = \chi(\sqrt{N}). \tag{4.8}$$

5. Random sequences

From now on we shall suppose, unless the context specifically denies it, that X is a sequence of identically distributed independent random variables, whose common distribution is continuous. This assumption of a continuous distribution conveniently ensures that, apart from a set of measure zero, no two terms of X are equal. Sets of measure zero will be ignored without further comment; so we can presume that $x_i \neq x_j$ for all $i \neq j$. Instead of supposing that the x_i were independent, we might have supposed them symmetrically dependent in Sparre Andersen's sense, or exchangeable (see [4], p. 225 for definitions of this); but the extra generality would be more apparent than real in the present context, and we shall not entertain it.

Clearly $\ell(X_N)$, $\ell'(X_N)$, and $\ell^*(X_N)$ are distribution-free random variables (that is to say, their distributions do not depend upon the common distribution of the terms of X); and, by symmetry, $\ell(X_N)$ and $\ell'(X_N)$ have the same distribution. What can we say about the distributions of $\ell(X_N)$ and $\ell^*(X_N)$, as Ulam asks, and what practical applications are there for the results? Ulam and Neighbor's results suggest that $N^{-1/2}\ell^*(X_N)$ converges to some random variable: what is it? We shall discover that a stronger asymptotic result holds.

THEOREM 4. *If X is a random sequence of the type described above, then $N^{-1/2}\ell(X_N)$ and $N^{-1/2}\ell^*(X_N)$ both converge in probability to an absolute constant c as $N \to \infty$. They also converge in* pth *absolute mean for any p satisfying $0 < p < \infty$.*

It seems very likely that these two random variables also converge with probability 1, but I cannot yet prove this conjecture.

To prove Theorem 4, we first assemble some remarks on subadditive stochastic processes and stochastic summation.

6. Subadditive and superadditive stochastic processes

Subadditive stochastic processes were first invented [15] to deal with time-dependent percolation processes; but they have other applications, one of which is to Ulam's problem. They also afford a generalization of renewal processes. Much remains to be done on the theory of these processes; and the original paper [15] lists a number of conjectures (the authors of a paper which does not furnish ample conjectures may be suspected, rightly or wrongly, of not working to the limits of their capabilities). With one exception, all these conjectures remain open: the exception is due to Kingman [18], who proved the ergodic theorem for subadditive processes.

A subadditive stochastic process is a family of real random variables $\{w_{r,s}(\omega)\}$ defined on a probability space (Ω, B, P) and indexed by a pair of nonnegative integers r, s such that $r \leqq s$. The process satisfies three postulates.

POSTULATE (i). *The process is* stationary *in the sense that its finite-dimensional distributions are the same as those of the shifted process*

$$w'_{r,s} = w'_{r,s}(\omega) = w_{r+1,s+1}(\omega). \tag{6.1}$$

POSTULATE (ii). *Each random variable of the process has finite expectation. The stationarity then ensures that this expectation depends only upon the difference of the indices r and s:*

$$E(w_{r,s}) = g_{s-r}, \text{ say}. \tag{6.2}$$

POSTULATE (iii). *For any* $\omega \in \Omega$ *and any three indices* r, s, t *such that* $r \leqq s \leqq t$, *we have the subadditive property*

$$w_{r,t}(\omega) \leqq w_{r,s}(\omega) + w_{s,t}(\omega). \tag{6.3}$$

From (6.2) and (6.3), we deduce

$$g_{u+v} \leqq g_u + g_v, \qquad u, v = 0, 1, 2 \cdots; \tag{6.4}$$

and hence, by a standard result on subadditive functions,

$$\lim_{t \to \infty} g_t/t = \inf_{t \geqq 1} g_t/t = c, \tag{6.5}$$

where c satisfies $-\infty \leqq c < \infty$. If g_t/t is bounded below then c is finite and is called the time constant of the process. In what follows, we always assume that the time constant exists.

A process $w_{r,s}$ is called superadditive if $-w_{r,s}$ is subadditive; that is, if the inequalities in (6.3) and (6.4) are reversed. Theorems for subadditive processes remain true for superadditive processes with obvious trivial modifications. If equality holds in (6.3) and (6.4), the theory reduces to the ordinary theory of additive processes. The postulates given above are those due to Kingman [18]; and are slightly more stringent than those in the original paper [15], where one-dimensional distributions were used in place of finite-dimensional ones in Postulate (i). The stricter requirement (i) is needed for the ergodic theorem, which runs as follows:

THEOREM 5. (Kingman) *Let* $w = \{w_{r,s}\}$ *be a subadditive stochastic process with a time constant* c; *and let* I *be the* σ*-field of events defined in terms of* w *and invariant under the shift* $w \to w'$. *Then as* $t \to \infty$, $t^{-1}w_{0,t}(\omega)$ *converges almost surely to a random variable* $W(\omega)$, *which can be expressed as a conditional expectation*

$$W(\omega) = \lim_{t \to \infty} t^{-1}E(w_{0,t}|I). \tag{6.6}$$

Moreover $W = c$ *almost surely when* I *consists only of events of probability* 0 *or* 1.

7. Stochastic summation

Many of the common procedures in the classical theory of summation [16] can be thought of in terms of discrete frequency distributions, though admittedly this is not the most usual way of looking at them. The purpose of summation is to assign a meaning to the statement

$$s_n = \sum_{i=0}^{n} u_i \to s = \sum_{i=0}^{\infty} u_i \qquad \text{as} \quad n \to \infty. \tag{7.1}$$

Consider a family of discrete probability distributions defined on the nonnegative integers, members of the family being indexed by some numerical parameter μ:

$$p_\mu = \{p_{0,\mu}, p_{1,\mu}, \cdots\}, \qquad p_{n,\mu} \geqq 0, \qquad \sum_{n=0}^{\infty} p_{n,\mu} = 1. \tag{7.2}$$

Quite commonly, μ is the mean of the distribution P_μ. Suppose the series

$$\sigma_\mu = \sum_{n=0}^{\infty} s_n p_{n,\mu} \tag{7.3}$$

converges in the ordinary sense; so σ_μ, the expectation of the partial sums s_n in (7.1), exists with respect to the distribution p_μ. If further σ_μ converges to a number s as $\mu \to \infty$, then we say (as the meaning to be assigned to (7.1)) that $\Sigma\, u_i$ is summable to s in the sense of p_μ; and we write (7.1) as

$$s = \sum_{i=0}^{\infty} u_i \qquad (p_\mu), \tag{7.4}$$

or as

$$s_n \to s \qquad (p_\mu). \tag{7.5}$$

For example, consider three famous particular cases of this procedure. First, if p_μ is the discrete uniform distribution

$$p_{n,\mu} = \begin{cases} N^{-1}, & n = 0, 1, \cdots, N-1, \\ 0, & n \geqq N, \end{cases} \qquad \mu = \tfrac{1}{2}(N-1), \tag{7.6}$$

we say that s is summable in the sense of Césaro, written $s_n \to s$ (C, 1). Second, if p_μ is the geometric distribution

$$p_{n,\mu} = p^n(1-p), \qquad \mu = p/(1-p), \tag{7.7}$$

we say that s is summable in the sense of Abel, written $s_n \to s$ (A). Third, if p_μ is the Poisson distribution

$$p_{n,\mu} = e^{-\mu}\mu^n/n!, \tag{7.8}$$

we say that s is summable in the sense of Borel, written $s_n \to s$ (B).

Besides being summable (p_μ), the series (7.4) or the sequence (7.5) may or may not be convergent in the ordinary sense. Two important classes of theorems deal with this situation. First, the so-called Abelian theorems assert (roughly speaking) that ordinarily convergent series (or sequences) are summable (p_μ). Second, the so-called Tauberian theorems assert the converse, provided that the terms u_i satisfy an additional condition (called a Tauberian condition) whose effect is to exclude the possibility of anomalous individual terms. For example, the Tauberian condition $u_n = O(n^{-1})$ suffices in order that Abel-summability should imply convergence.

Now the classical theory of summation deals with real variables u_i or s_n; but there is no reason why it should not be extended to random variables. I do not know whether the literature contains a systematic extension along these lines, though certainly there are isolated cases dealing mainly with Césaro-summation of random variables: if no such systematic extension exists, it could perhaps provide a straightforward topic for a doctoral dissertation. Here I have raised the subject because in the next section I shall use a Tauberian argument on Borel-summation of random variables.

8. Asymptotic behavior in Ulam's problem

Consider all possible subsequences of length n contained in X_N, and let $m_{n,N}$ denote the expected number of these which are monotone.

THEOREM 6. *If* $n = \chi(e\sqrt{N}) + t$ *and* $t \geqq 0$, *then*

$$P(\ell_N^* \geqq n) \leqq m_{n,N} = \frac{2}{n!}\binom{N}{n} \leqq \frac{e^{-2t}}{\pi\sqrt{N}}. \tag{8.1}$$

If ρ_k is the probability that exactly k of these subsequences are monotone, then

$$P(\ell_N^* \geqq n) = \sum_{k \geqq 1} \rho_k \leqq \sum_{k \geqq 0} k\rho_k = m_{n,N} = \frac{2}{n!}\binom{N}{n}; \tag{8.2}$$

for there are $\binom{N}{n}$ such subsequences and $2/n!$ is the probability that any specified one of them will be monotone. Also, since $n \geqq \chi(e\sqrt{N})$,

$$\frac{m_{n+1,N}}{m_{n,N}} = \frac{N-n}{(n+1)^2} \leqq \frac{N}{(e\sqrt{N})^2} = e^{-2}; \tag{8.3}$$

and hence

$$m_{n,N} \leqq e^{-2t} m_{\chi(e\sqrt{N}),N}. \tag{8.4}$$

Hence, to complete the proof of Theorem 6, we need only establish that

$$m_{n,N} \leqq 1/\pi\sqrt{N} \tag{8.5}$$

holds in the special case

$$n = \chi(e\sqrt{N}). \tag{8.6}$$

Now if $N < 8$, we find $n > N$ and therefore $m_{n,N} = 0$ and (8.5) is trivially true. For the cases $8 \leqq N < 16$, calculation yields

(8.7)

N:	8	9	10	11	12	13	14	15
n:	8	9	9	10	10	10	11	11
$11!m_{n,N}$:	990	110	1100	121	726	3146	364	1365.

Thus in any of these eight cases, we have

$$m_{n,N} \leqq 3146/11! \leqq 1/4\pi < 1/\pi\sqrt{N}, \qquad 8 \leqq N < 16. \tag{8.8}$$

So it remains to prove (8.5) for $N \geqq 16$. Using Stirling's formula in the form

$$\log a! = (a + \tfrac{1}{2}) \log a - a + \tfrac{1}{2} \log (2\pi) + \frac{\theta}{12a}, \qquad 0 < \theta < 1, \tag{8.9}$$

we deduce

$$\begin{aligned}(8.10)\qquad \log m_{n,N} \leqq\ & \log 2 + (N + \tfrac{1}{2}) \log N - N + \tfrac{1}{2} \log (2\pi) + (12N)^{-1} \\ & - (2n + 1) \log n + 2n - \log (2\pi) \\ & - (N - n + \tfrac{1}{2}) \log (N - n) + (N - n) - \tfrac{1}{2} \log (2\pi) \\ \leqq\ & n + n \log \frac{N}{n^2} - (N - n + \tfrac{1}{2}) \log\left(1 - \frac{n}{N}\right) - \log (\pi n) + \frac{1}{12N}.\end{aligned}$$

But $N/n^2 \leqq e^{-2}$; and also

$$(8.11)\qquad - (N - n) \log\left(1 - \frac{n}{N}\right) \leqq (N - n) \sum_{r=1}^{\infty} \left(\frac{n}{N}\right)^r = n.$$

Hence

$$(8.12)\qquad \log m_{n,N} \leqq - \tfrac{1}{2} \log\left(1 - \frac{n}{N}\right) - \log (\pi n) + (12N)^{-1}.$$

With $N \geqq 16$, we have

$$(8.13)\qquad \frac{n}{N} \leqq \frac{e\sqrt{N} + 1}{N} = \frac{1}{\sqrt{N}}\left(e + \frac{1}{\sqrt{N}}\right) \leqq \tfrac{1}{4}(e + \tfrac{1}{4}):$$

and therefore

$$\begin{aligned}(8.14)\qquad \log m_{n,N} &\leqq - \log (\pi n) - \tfrac{1}{2} \log \frac{15 - 4e}{16} + \frac{1}{192} \\ &\leqq - \log (\pi e\sqrt{N}) - \tfrac{1}{2} \log \tfrac{1}{4} + \frac{1}{192} \leqq - \log (\pi\sqrt{N}).\end{aligned}$$

This proves (8.5), as required.

Theorem 6 shows that the upper tail of the distribution of $\ell^*_N/\sqrt{N}$, that is to say from e upwards, has probability $O(N^{-1/2})$ as $N \to \infty$; and likewise, because $\Sigma_{t=0}^{\infty} t^p e^{-2t}$ converges, the contribution from this tail to any pth absolute moment of $\ell^*_N/\sqrt{N}$ is $O(N^{-(p+1)/2})$. Thus in proving Theorem 4, there is no essential loss of generality in treating $\ell^*_N/\sqrt{N}$ as though it were a bounded random variable, restricted to the closed interval $[0, e]$. Accordingly, convergence in pth absolute mean will follow if we can prove convergence in probability. The same remarks apply to $\ell_N/\sqrt{N}$ because $\ell_N \leqq \ell^*_N$. Lastly, ℓ_N and ℓ'_N have the same distribution; so $\ell^*_N/\sqrt{N} = \max (\ell_N/\sqrt{N}, \ell'_N/\sqrt{N})$ will converge in probability to c if $\ell_N/\sqrt{N}$ does.

Thus it remains to prove that

$$(8.15)\qquad \ell_N \sim c\sqrt{N}$$

in probability as $N \to \infty$. There are, I suppose, three stages in solving a problem.

The first stage is to get a clear idea of the essence of the problem, and to clear away the minor irrelevancies. I have just done this for Theorem 4: equation (8.15) contains the essence of Theorem 4, and the statements about convergence in pth absolute mean are minor irrelevancies. This concentration upon the main issue may easily, and here does, simplify things: the distinction between ℓ_N^* and ℓ_N is comparatively unimportant, but ℓ_N is an easier quantity to handle. The second stage is to conceive the basic ideas, on which the proof will rest; and the third stage is to work out the details of the proof. This third stage is normally quite easy, just a matter of craftsmanship. The second stage is by far the most difficult, since it requires a certain power of imagination coupled with a mathematical alertness. My own approach, for what it is worth, is to run quickly through a catalogue of available mathematical tools: such a catalogue is liable to be quite short. Here (8.15) is a fairly hard result to get at; and most of the familiar tools (for example, law of large numbers, Markov processes, renewal theory, and so forth) can be eliminated as insufficiently powerful tools for the job. The only two tools, which seemed to me to be strong enough for the task, were subadditive processes and submartingales. Being more familiar with the former, I started with them; and since they worked, I did not pursue the other alternative. However, devotees of submartingales will doubtless wish to explore the latter possibility. In considering each tool in the catalogue, one has to envisage various different ways of using it. One needs to think of the problem upside down and inside out, as it were, and to entertain unusual ways of handling the tool. This is where the alertness comes in; for otherwise one may miss the elusive idea that does the trick. The situation is rather like playing a game of chess: one follows certain strategies, but one always has to be alert for the position that conceals a winning combination. In the present case, the main difficulty to surmount is that subadditive processes are essentially linear, namely

$$w_{0,t} \sim ct, \qquad (8.16)$$

whereas (8.15) is nonlinear. How does one introduce the square root? There is a clue in the fact that it is a *square* root: why not make use of the geometrical properties of a square? Readers, who would like to evaluate their sense of mathematical alertness, may care to pause at this point and ask themselves what is going to happen next, *given* that subadditive processes, Borel-summation, and the geometry of a square are the combination of ideas that will prove (8.15).

To return to the chess analogy, you may imagine that you are faced with a chess columnist's game position, for which you are *told* that a winning combination exists and, moreover, can be achieved by a queen sacrifice followed by a discovered check and a pawn promotion. Of course, all that sort of information makes things much easier than it would have been if you had actually been playing the game itself and had had to make yourself aware of the existence and nature of the combination. If you wish to think out for yourself how (8.15) may be proved, do not turn over the page yet—for to do so would give the game away too soon.

Consider a Poisson process of unit density in the Euclidean plane with a coordinate system (ξ, η). For nonnegative integers r, s $(r \leqq s)$ let $S_{r,s}$ be the square

$$r \leqq \xi < s, \qquad r \leqq \eta < s, \tag{8.17}$$

with the convention that this square is the empty set if $r = s$. We say that a set of points (ξ_i, η_i), $i = 1, 2, \cdots, k$, in the plane form a *chain* if they can be connected together by a path that proceeds in a northeasterly direction, where "northeast" means any direction between north and east inclusive: that is to say, if and only if

$$\xi_1 \leqq \xi_2 \leqq \cdots \leqq \xi_k \quad \text{and} \quad \eta_1 \leqq \eta_2 \leqq \cdots \leqq \eta_k. \tag{8.18}$$

The length of a chain is defined to be the number of points in the chain. Now consider the points of the Poisson process, which fall in the square $S_{r,s}$, and let $w_{r,s}$ be the length of a longest chain which can be formed from some subset of these points. Here $w_{r,s} = 0$ if $S_{r,s}$ is empty or contains no points of the Poisson process.

Since the squares $S_{r,s}$ and $S_{s,t}$ are contained in $S_{r,t}$, and any point in $S_{s,t}$ is to the northeast of any point in $S_{r,s}$, we have

$$w_{r,t} \geqq w_{r,s} + w_{s,t}, \qquad r \leqq s \leqq t. \tag{8.19}$$

The finite-dimensional distributions of the process $\{w_{r,s}\}$ are invariant under the shift $w_{r,s} \to w'_{r,s} = w_{r+1,s+1}$ by the spatial homogeneity of the Poisson process.

Let τ be the number of points of the Poisson process in $S_{0,t}$; and suppose that these points have coordinates (ξ_i, η_i) where $i = 1, 2, \cdots, \tau$. Let us also order these points so that

$$\xi_1 < \xi_2 < \cdots < \xi_\tau. \tag{8.20}$$

Here we may ignore the possibility that any two ξ or any two η are equal, for this event has zero probability. The corresponding sequence $\{\eta_1, \eta_2, \cdots, \eta_\tau\}$ will be a sequence of identically distributed independent random variables; and $w_{0,t}$ will be the length of the longest ascending subsequence in $\{\eta_1, \eta_2, \cdots, \eta_\tau\}$. So $w_{0,t}$ has the same distribution as ℓ_τ. The random variable τ has a Poisson distribution with parameter t^2, the area of $S_{0,t}$; and hence, using Theorem 6, we have

$$\begin{aligned} Ew_{0,t} &= \sum_{N=0}^{\infty} e^{-t^2} \frac{t^{2N}}{N!} E(\ell_N) \\ &= O\left\{ \sum_{N=0}^{\infty} e^{-t^2} \frac{t^{2N}}{N!} \sqrt{N} \right\} = O(t) \qquad \text{as} \quad t \to \infty. \end{aligned} \tag{8.21}$$

So $t^{-1}Ew_{0,t}$ is bounded. We have now verified all three postulates in Section 6, and we can conclude that $\{w_{r,s}\}$ is a superadditive stochastic process. By Theorem 5, there exists a random variable W such that

$$(8.22)\qquad t^{-1}w_{0,t} \to W \quad \text{almost surely as} \quad t \to \infty.$$

The points of the Poisson process which lie in the unit square $r \leqq \xi < r + 1$, $s \leqq \eta < s + 1$ can be generated from a random variable ν (having a Poisson distribution with parameter 1) and an infinite collection of independent observations from a uniform distribution on $[0, 1)$ from which we utilize the first 2ν observations to yield the coordinates of ν points in the square. We write $\zeta_{r,s}$ for the collection of all random variables associated in this way with this square. The family $\{\zeta_{r,s}; r, s = 0, 1, \cdots\}$ is a family of mutually independent collections; and $w_{r,s}$ can be written as a fixed function

$$(8.23)\qquad w_{r,s} = F\{\zeta_{r,s}, \zeta_{r+1,s}, \cdots, \zeta_{s+1,r}, \cdots\}.$$

In particular, $w_{r,s}$ is independent of those $\zeta_{\rho,\sigma}$ with $\rho < r$ and $\sigma < s$. So the invariant σ-field of the $w_{r,s}$-events can be embedded in the remote σ-field of the $\zeta_{r,s}$; and hence consists only of events of probability 0 or 1. Thus W is almost surely a constant c; and we conclude that

$$(8.24)\qquad t^{-1}w_{0,t} \to c \quad \text{almost surely as} \quad t \to \infty.$$

It only remains to find a Tauberian argument for unscrambling the Borel-summation induced by the Poisson process. To this end we extend the definition of $w_{0,t}$: for any real $t \geqq 0$, no longer an integer necessarily, we say that $w_{0,t}$ is the length of the longest chain amongst the points of the Poisson process in the square $0 \leqq \zeta < t, 0 \leqq \eta < t$. For each given realization of the Poisson process, $w_{0,t}$ is a nondecreasing function of t: this fact serves as the required Tauberian condition. In the first place it ensures that (8.24) remains true for any sequence of real t tending to infinity, and therefore for any sequence of real random variables t tending to infinity with probability 1. Next we define $t(N)$ by the requirement that it is the smallest value of t such that the square $0 \leqq \xi < t, 0 \leqq \eta < t$ shall contain exactly N points of the Poisson process. Here $N = 1, 2, \cdots$; and for each given N, the distribution of $w_{0,t(N)}$ is the same as the distribution of ℓ_N in Ulam's problem. Moreover, from the properties of the Poisson process

$$(8.25)\qquad t(N)/\sqrt{N} \to 1 \quad \text{almost surely as} \quad N \to \infty.$$

Putting $t = t(N)$ in (8.24) we deduce that

$$(8.26)\qquad w_{0,t(N)}/\sqrt{N} \to c \quad \text{almost surely as} \quad N \to \infty.$$

Although we have almost sure convergence in (8.26), we have not proved almost sure convergence in Theorem 4: the reason is that $w_{0,t(N)}$ in (8.26) is not associated with a sequence X in the way in which ℓ_N is in the statement of Theorem 4. It merely has the same distribution as ℓ_N for each *given* N. Therefore we have only proved convergence in probability in Theorem 4.

THEOREM 7. *The constant c in Theorem* 4 *satisfies*

$$(8.27)\qquad \tfrac{1}{2}\pi \leqq c \leqq e.$$

Consider the points of the Poisson process (with unit parameter) and a square of area N. For any point P in the square, let $Q(P)$ denote the point of the Poisson process which is northeast of P and as close to P as possible. Let Q_0 be the southwest corner of the square, and from the sequence

$$Q_{i+1} = Q(Q_i), \qquad i = 0, 1, 2, \cdots. \tag{8.28}$$

The expected value of the horizontal (or vertical) projection of $Q_i Q_{i+1}$ is

$$\int_0^\infty dr \int_0^{\pi/2} d\theta (re^{-\pi r^2/4})(r \cos \theta) = \frac{2}{\pi}. \tag{8.29}$$

These projected lengths are independent; and so the strong law of large numbers shows that (with probability 1 as $N \to \infty$) $\frac{1}{2}\pi\sqrt{N} + o(\sqrt{N})$ terms of (8.28) can be formed before reaching the opposite boundary of the square. The sequence (8.28) provides a chain, not longer than a longest chain in the square. This proves the lower bound in (8.27); and the upper bound follows at once from Theorem 6. Professor Kingman has remarked to me that, if in the third line of the proof of Theorem 7 we interpret the word "close" in terms of distance measured parallel to the diagonal of the square, we get a different integral in (8.29) with a value $(\pi/8)^{1/2}$. Thus the lower bound in (8.27) can be raised from $\frac{1}{2}\pi = 1.57 \cdots$ to $(8/\pi)^{1/2} = 1.59 \cdots$. Professor Blackwell has also obtained this result independently.

9. Monte Carlo methods

In this section I shall describe how to study the behavior of ℓ_N by a Monte Carlo method, called dummy truncation. This device, originally introduced many years ago to deal with percolation processes [12], is really a very simple idea. In studying a quantity, such as $L_N = N^{-1/2}\ell_N$, which converges in probability as $N \to \infty$, it may be better to spend a given amount of computing time on a small number of samples with large N rather than a large number of samples with small N. Too small a value of N entails the risk of not reaching the region of asymptotic behavior; but a small sample size with large N need not, on the other hand, prejudice the accuracy of the estimation because the sampling variance of L_N becomes small as N increases on account of the convergence. Ulam's account of Neighbor's Monte Carlo work does not mention the sample size (though one may guess it was large since a computer was used), but Ulam does state that the values of N were small ($N \leq 101$); and we shall see presently that his N are all much too small to represent the true asymptotic behavior. How does one know when the asymptotic region is reached? There is no panacea, of course; but a reasonable procedure is to study successive values of N until L_N appears to settle down to some stable value. Thus we regard the whole vector

$$\{\ell(X_1), \ell(X_2), \cdots, \ell(X_N), \cdots\}, \tag{9.1}$$

as a single observation of the Monte Carlo sample, and the different observations of the sample come from taking different sequences X. The sample size will be small if we only consider a small number of different X. In practice we have to terminate (9.1) at some value of N, this value being chosen when $L(X_N) = N^{-1/2}\ell(X_N)$ reaches stability. Here we have to be cautious: since X is common to all terms in (9.1), the terms in (9.1) are highly correlated. Therefore the stability of $L(X_N)$ for each individual X is not enough: we must also check that the stable values for different X agree with each other.

The plan will be impractical unless we have an efficient computing algorithm for generating the successive coordinates of (9.1). To this end, suppose that

$$X = \{x_1, x_2, \cdots\} \tag{9.2}$$

consists of independent identically distributed observations x_i from a uniform rectangular distribution on the interval $[0, 1]$; and introduce a real variable x, called the dummy truncator, which can take any value in $[0, 1]$. We define X_N^x to be the subsequence of

$$X_N = \{x_1, x_2, \cdots, x_N\} \tag{9.3}$$

obtained by deleting from X_N all $x_i > x$. We write, as usual, $\ell(X_N^x)$ for the length of the longest ascending subsequence of X_N^x; and we now regard $\ell(X_N^x)$ as a function of x for each N. Of course

$$\ell(X_N) = \ell(X_N^1); \tag{9.4}$$

so we can recover (9.1) if we have

$$\{\ell(X_1^x), \ell(X_2^x), \cdots, \ell(X_N^x), \cdots\} \tag{9.5}$$

available in the computer. However, this is more than we need store in the computer. To see this, we note that $\ell(X_N^x)$ is an integer step function of x, satisfying the recurrence relation

$$\ell(X_{N+1}^x) = \begin{cases} \ell(X_N^x\}, & 0 \leqq x < x_{N+1} \\ \max\{\ell(X_N^x), 1 + \ell(X_N^{x_{N+1}})\}, & x_{N+1} \leqq x \leqq 1. \end{cases} \tag{9.6}$$

This recurrence relation starts from

$$\ell(X_1^x) = \begin{cases} 0, & 0 \leqq x < x_1 \\ 1, & x_1 \leqq x \leqq 1. \end{cases} \tag{9.7}$$

Moreover, $\ell(X_N^x)$ is completely specified by a statement of the positions of its steps: suppose these occur at $x = y_{i,N}$, $i = 1, 2, \cdots, I(N)$, where

$$y_{1,N} < y_{2,N} < \cdots < y_{I(N),N}. \tag{9.8}$$

From (9.6), we see that

$$y_{1,N+1} < y_{2,N+1} < \cdots < y_{I(N+1),N+1} \tag{9.9}$$

is obtained from (9.8) by adding x_{N+1} to the end of (9.8) if $x_{N+1} > y_{I(N),N}$, and otherwise by substituting x_{N+1} in place of the least $y_{i,N} \geqq x_{N+1}$. The recurrence from (9.8) to (9.9) starts from

$$y_{1,1} = x_1. \tag{9.10}$$

Thus all we need do is to store (9.8) in the computer and successively update it to (9.9).

This procedure looks formidable at first sight: but a numerical example will show that it is really very simple. Suppose that

$$X = \{0.23, 0.47, 0.14, 0.22, 0.96, 0.83, \cdots\} \tag{9.11}$$

Then

$$\begin{aligned} \{y_{1,1}\} &= \{0.23\} \\ \{y_{1,2}, y_{2,2}\} &= \{0.23, 0.47\} \\ \{y_{1,3}, y_{2,3}\} &= \{0.14, 0.47\} \\ \{y_{1,4}, y_{2,4}\} &= \{0.14, 0.22\} \\ \{y_{1,5}, y_{2,5}, y_{3,5}\} &= \{0.14, 0.22, 0.96\} \\ \{y_{1,6}, y_{2,6}, y_{3,6}\} &= \{0.14, 0.22, 0.83\} \\ &\text{-------------} \end{aligned} \tag{9.12}$$

For hand computing, this can be achieved by writing the terms of (9.11) in rows (pigeonholes, rather like those in the proof of Theorem 2 but with ascending and descending roles reversed); and ℓ_N will be the number of rows used, while the y are the last entries in each row. Successive appearances of this tableau will look like:

0.23	→	0.23	→	0.23, 0.14	→	0.23, 0.14	→	0.23, 0.14	→	0.23, 0.14
		0.47		0.47		0.47, 0.22		0.47, 0.22		0.47, 0.22
								0.96		0.96, 0.83

Since we have

$$\ell(X_N) = I(N), \tag{9.13}$$

we shall generate (9.1) if we generate X from a sequence of pseudorandom numbers, which we then sort in an overwritten list (9.8), the length of the list at any instant n yielding $\ell(X_n)$. From Theorem 4, the length of the final list will be about $c\sqrt{N}$. Hence to generate (9.1) as far as N, the storage requirement will be approximately $c\sqrt{N}$ and the computing time will be proportional to $\Sigma_{n=1}^{N} c\sqrt{n} = \frac{2}{3} cN^{3/2}$. This is not excessive even for values of N as large as a million. (Programming experts will see that the foregoing computing time can be considerably shortened by appropriate block addressing and address modifiers; but I shall not go into that here.)

Dr. D. C. Handscomb and Mrs. L. Hayes were kind enough to program the foregoing algorithm on the computer at Oxford. The program was written to accept any value of $N \leqq 10^6$; but due to pressure of time (I was due to fly to Berkeley at the end of that week), we only ran the program up to $N = 10^4$, we only printed out values of $\ell(X_N)$ for $N = 100, 400, 1600, 4900, 10000$; and we only did this for 10 different sequences X. This, of course, is a mere sketch of a Monte Carlo calculation; and a full-sized calculation ought to be undertaken in due course of time. However, even these sketchy calculations with a sample size of 10 provide some interesting results. In Table I, the 10 lines correspond to the 10 different sequences X, the quantities tabulated being $\ell(X_N)$.

TABLE I

MONTE CARLO OBSERVATIONS OF $\ell(X_N)$

$N = 100$	$N = 400$	$N = 1600$	$N = 4900$	$N = 10000$
20	38	76	133	198
18	40	76	132	197
17	40	75	132	197
16	35	74	134	198
17	35	74	135	198
18	39	76	132	197
18	37	74	136	195
16	35	74	134	198
19	39	76	132	197
17	39	75	132	197

The consistency between sequences is impressive: indeed I have an uneasy feeling that it is fortuitously too good, and that a larger sample would reveal more scatter. However, for what these data are worth, we get the following

TABLE II

MONTE CARLO ESTIMATES

N	$E(\ell_N/\sqrt{N})$	Var ℓ_N
100	1.76 ± 0.04	1.6
400	1.88 ± 0.03	3.8
1600	1.88 ± 0.007	0.8
4900	1.90 ± 0.005	1.5
10000	1.97 ± 0.003	0.8

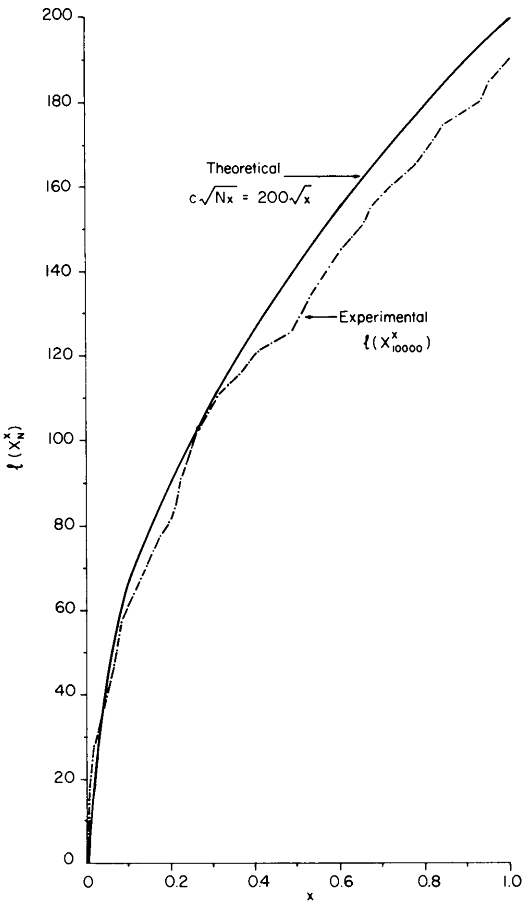

FIGURE 1
Monte Carlo experiment for $\ell(X_N^x)$ with $N = 10000$.

estimates. The $\pm$ entries in the second column are standard errors: these and the estimates of variance in the third column are certainly ragged and seem to me to be suspiciously small. Better estimates must await more extensive Monte Carlo calculations. The mean length 17.6 ± 0.4 for $N = 100$ agrees with the Ulam-Neighbor figure of 17.85 for $N = 101$ (his figures were for the slightly larger quantity ℓ^* in place of ℓ); but the second column suggests that $\ell_N/\sqrt{N}$ does not come close to c until N is substantially larger, say $N = 10000$ or more. The value of c seems to be near 2. Figures 1 and 2 show the results of an eleventh observation. Figure 1 gives a graph of $\ell(X_N^x)$ for $N = 10000$ and $0 \leqq x \leqq 1$. Figure 2 shows the convergence of $N^{-1/2}\ell(X_N)$ for $N = 1, 2, \cdots, 10000$. This eleventh observation is markedly smaller than the other 10 observations.

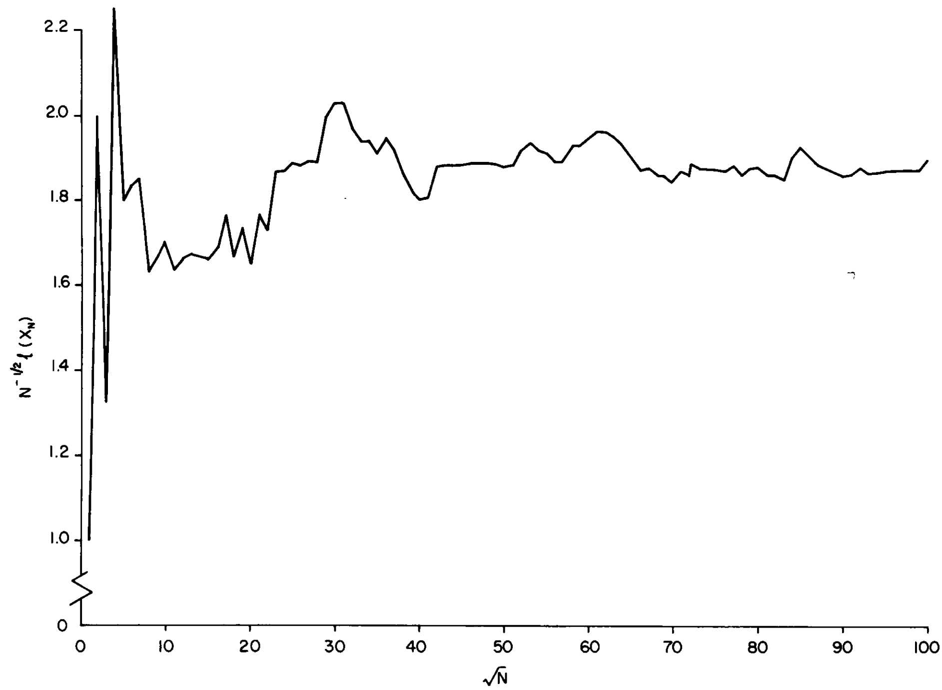

FIGURE 2
Monte Carlo experiment for $N^{-1/2}\ell(X_N)$ with $N = 1, 2, \cdots, 10000$.

10. Attack on c: first method

The arguments in the first eight sections of this paper are rigorous; or rather (since "rigorous" is a subjective term), people, finding the arguments nonrigorous, should at least have little difficulty in modifying them to accord with their own personal standards. However, as usual the price paid for rigor is conclusions which are rather insipid and general, which merely assert the existence of limits or the qualitative behavior of functions. This and the next two sections, on the other hand, deal with the harder problem of assigning a numerical value to the constant c, whose mere existence was established by Theorem 4. Here I jettison rigor without compunction, because a premature attempt to retain it would quite simply halt the work. Of course, I do not mean that rigor should be scorned for its own sake, but merely that the initial investigation is best done without the shackles of rigor and I have not yet got beyond this initial stage. A research worker needs to be able to think nonrigorously in order to get off the ground in the first place, and this ability is in fact rather more difficult to achieve than rigorous thought; for it requires a rather sophisticated sense of judgment over the adequacy of approximations, the plausibility of the reasoning, and the prospects for ultimately making it watertight. One should try to make many different attacks upon a difficult problem—wave after wave of attack if possible, from all sorts of angles, to test the weak points of the defenses and reveal the footholds from which a final assault may be launched. This preliminary reconnoitering and skirmishing should be highly informal, though not without a perceptive eye for routes able to support the weight of later formality. I shall try, rather sketchily, to illustrate this by presenting three separate nonrigorous determinantions of c. The first determination has a superficial attractiveness; but it is, I believe, thoroughly disreputable. The second and third arguments look progressively wilder but are, I think, more promising. But these assessments are matters of taste and experience, and the reader must make up his own mind about them.

Consider, as in the proof of Theorem 7, a Poisson process with unit parameter and a square of area N. Let $P_1, P_2, \cdots, P_\ell$ be the points of a longest chain in the square drawn from the points of the Poisson process. Here

$$\ell = \ell_N = c\sqrt{N}, \tag{10.1}$$

with (of course) a suitable nonrigorous interpretation upon the equality signs in (10.1). The average horizontal (or vertical) displacement between two successive points P_i, P_{i+1} is $1/c$ since the square has side $\sqrt{N}$; and hence the area of a rectangle R, having sides parallel to the square and opposite vertices at P_i and P_{i+2} (*not* P_{i+1}) is about $4/c^2$. However, the rectangle R may be expected to contain just one point (P_{i+1}) of the Poisson process (if it contained two or more points there would be at least a 50 per cent chance of two or more points of the chain in R). The expected number of Poisson points in an area $4/c^2$ is $4/c^2$. Hence

$$4/c^2 = 1, \tag{10.2}$$

which gives $c = 2$.

This looks like the right value of c, and at first glance the method has a pleasant appearance of simplicity. On closer examination, however, it becomes far less attractive. There is in fact a vicious circularity about the argument; and this is best exhibited by considering instead the rectangle R' with opposite vertices at P_i and P_{i+1}. The area of R' is about $1/c^2$, and R' certainly contains no points of the Poisson process. But $1/c^2$ is nothing like zero. The trouble arises because R' depends upon the properties of the chain, and hence upon the Poisson process. There seems little prospect of mending this first method.

This is a pity, because an extension of this method would have yielded not merely c, but also the values of

$$\lambda(N) = E(\ell_N) \tag{10.3}$$

for finite N. To pass quickly over this, consider the rectangle R'' with opposite vertices at P_i and P_{i+k}. This has area k^2/c^2; and so the probability that there are n Poisson points in R'' is

$$e^{-k^2/c^2}(k^2/c^2)^n/n! \tag{10.4}$$

The expected length of longest chain from these n points in R'' is $\lambda(n)$ from (10.3); and R'' is known to contain a longest chain of length $k - 1$. Hence

$$\sum_{n=0}^{\infty} e^{-k^2/c^2} \frac{(k^2/c^2)^n}{n!} \lambda(n) = k - 1. \tag{10.5}$$

There are, of course, obvious queries over summing from $n = 0$ rather than from $n = k$, say; but let us ignore them.

The left side of (10.5) is what might be called a Borel transform, by analogy with Section 7. How does one invert a Borel transform? Putting

$$k^2/c^2 = \theta, \tag{10.6}$$

we get

$$\sum_{n=0}^{\infty} e^{-\theta} \frac{\theta^n}{n!} \lambda(n) = c\theta^{1/2} - 1; \tag{10.7}$$

and hence taking Laplace transforms with respect to θ, we have

$$\sum_{n=0}^{\infty} \frac{\lambda(n)}{(s + 1)^n} = \frac{c\sqrt{\pi}}{2s^{3/2}} - \frac{1}{s}. \tag{10.8}$$

This is a trifle nonsensical: the right side has a branch point, and it becomes negative when $s \to 0$ while the left side remains obstinately positive. A source of this discrepancy is that, in taking Laplace transforms, we have treated θ as a continuous variable from 0 to ∞ although (10.6) restricts it to discrete values. There are infinitely many continuous functions of θ which agree with the right side of (10.7) for these discrete values: which of these functions can we choose to make the right side of (10.8) positive?

However, to continue, we put $t = 1/(s + 1)$ in (10.8) and expand the right side in powers of t, obtaining

$$\sum_{n=0}^{\infty} \lambda(n)t^n = c \sum_{n=0}^{\infty} \frac{(n + \frac{1}{2})!}{n!} t^{n+3/2} - \sum_{n=1}^{\infty} t^n. \tag{10.9}$$

In this we want to equate coefficients. In ordinary circumstances we could replace a series like $\Sigma\, a_n t^{n+1}$ by $\Sigma\, a_{n-1} t^n$; and we extend this principle of shifting n by an integer to nonintegral shifts. This yields

$$\lambda(n) = c \frac{(n - 1)!}{(n - \frac{3}{2})!} - 1. \tag{10.10}$$

For large n, this gives $\lambda(n) \sim c\sqrt{n}$ as it should. For small n, we insert $c = 2$ from (10.2) and calculate the right side of (10.10). Table III compares these results with the exact values of $\lambda(n)$ obtained later in Section 17.

TABLE III

VALUES OF $\lambda(N)$

N	Formula (10.10)	Exact $\lambda(N)$
2	1.27	1.50
3	2.01	2.00
4	2.61	2.42
5	3.13	2.79
6	3.59	3.14
7	4.00	3.47
8	4.39	3.77
9	4.75	4.06

The agreement is only moderate, perhaps all that could have been expected from an outlandish calculation like this one. However, some concealed ideas may possibly lurk in the foregoing, and perhaps they can be discovered by a little gentle cooking of the mathematics and a proper disregard for the occasional whiff of burning.

11. Attack on c: second method

It is more or less true (in some sense) that

$$\ell(X_N) = c\sqrt{N} + o(\sqrt{N}). \tag{11.1}$$

If the error term $o(\sqrt{N})$ behaves smoothly enough, then

$$E\{\ell(X_{N+1}) - \ell(X_N)|\ell(X_N)\} \sim c\sqrt{(N+1)} - c\sqrt{N} \sim \frac{c}{2\sqrt{N}}. \tag{11.2}$$

Now consider a square S with N points uniformly distributed in it; and consider adding one more point to S, this extra point being also uniformly distributed over S. We look at the conditional situation, given the positions of the original N points. The extra point can only increase a longest chain of length $\ell(X_N)$ by 1. Hence the expected conditional increase, namely (11.2), is equal to the probability that the new point will cause an increase of 1. Thus the area of the region in S, in which the new point will cause an increase, is $c/2\sqrt{N}$.

Consider a longest chain $P_1, P_2, \cdots, P_\ell$ from the original N points. Take P_0 and $P_{\ell+1}$ to be the southwest and northeast corners of S. Let $R_i, i = 0, 1, \cdots, \ell$, be the rectangle with opposite vertices P_i and P_{i+1}. The new point will cause an increase of 1 to *this* chain if and only if it falls into one of the rectangles R_0, $R_1, \cdots, R_\ell$. If these rectangles were squares, all of equal size, their total area would be $(\ell + 1)/(\ell + 1)^2 \sim 1/c\sqrt{N}$. Actually their total area will be larger than this (actually about twice as large), because they are not of equal size. There will be a similar chain of rectangles from any other longest chain $P'_1, P'_2, \cdots, P'_\ell$; and the rectangles from one chain will overlap those of another chain. Let us hope that the overlapping compensates more or less for the underestimation of the total area associated with an individual chain. Hence the average number of points which form an ith point of some longest chain is

$$(c/2\sqrt{N})/(1/c\sqrt{N}) = \tfrac{1}{2}c^2. \tag{11.3}$$

This holds for each given value of i; and therefore (ignoring the distinction between arithmetic and geometric means) the total number of longest chains is about

$$(\tfrac{1}{2}c^2)^\ell \sim (\tfrac{1}{2}c^2)^{c\sqrt{N}}. \tag{11.4}$$

On the other hand from Theorem 6, the expected number of increasing subsequences of length $c\sqrt{N}$ is

$$\frac{1}{(c\sqrt{N})!}\binom{N}{c\sqrt{N}} \sim \left(\frac{e}{c}\right)^{2c\sqrt{N}}. \tag{11.5}$$

Equating (11.4) and (11.5) we get

$$\tfrac{1}{2}c^2 = (e/c)^2 \tag{11.6}$$

which leads to

$$c = 2^{1/4}e^{1/2} = 1.961\ldots. \tag{11.7}$$

This method is admittedly very rough and ready: the saving grace is the fourth root at the end of the calculation, which reduces any relative error by a factor of 4. It is the only method I have been able to invent which involves a final fourth root: the other two methods presented here, as well as further methods which I shall not mention, end by taking a square root. The suggestion accordingly is that one should try to look for a method which ends by taking an arbitrarily high root, thus swamping any approximations in the early part of the calculation (Littlewood's "high indices" principle).

At first sight it seems odd that the number of longest chains is as large as (11.5). The reason becomes apparent when one considers three successive terms $y < y' < y''$ in a longest ascending subsequence of X. There is a reasonable chance that X contains another term z, say, such that $y < z < y''$ while $z > y'$. Thus the term y' may be replaced by z. This sort of substitution can take place along the whole length of the chain; and hence the number of longest chains will increase exponentially with the length of a longest chain. Actually I can prove rigorously that the expected number of longest chains is at least

$$(e/c)^{2c\sqrt{N}+o(\sqrt{N})}. \tag{11.8}$$

The reason why I have to be content with this as a lower bound is that there is a small probability of the length of a longest chain being less than $c\sqrt{N}$; if a longest chain is shorter than $c\sqrt{N}$, the number of such chains increases exponentially over and above (11.8). This very large increase in numbers may override the small probability.

12. Attack on c: third method

This method is based on the principle that the only good Monte Carlo method is a dead one: to wit, having done the appropriate transformation to prepare a problem for Monte Carlo sampling (as by dummy truncation in Section 9), one then ought to abandon the proposed Monte Carlo sampling and instead investigate the transformed problem analytically. (If the analysis appears intractable, the transformation is not an adequate preparation for Monte Carlo sampling!)

The analytical idea adopted here is to calculate the area under the curve $\ell(X_N^x)$, defined in (9.6) and regarded as a function of x. Since the number of terms in the sequence X_N^x will be asymptotically Nx for large N, we have

$$\ell(X_N^x) = c\sqrt{Nx} + o(\sqrt{N}). \tag{12.1}$$

The area under this curve is

$$A = \int_0^1 \ell(X_N^x)\,dx. \tag{12.2}$$

We shall calculate A in two ways, one of which expresses it as a multiple of $1/c$. By equating these two results, we then determine c. The second calculation, leading to a multiple of $1/c$, might seem to conflict with the dogma that integration is a linear operation: the moral is that in searching for methods of calculating scale factors one should not be blinded by the dogmas of functional analysis.

The first calculation of A is straightforward: from (12.1) and (12.2) we have

$$\mathrm{A} = \tfrac{2}{3}c\sqrt{N} + o(\sqrt{N}). \tag{12.3}$$

For the second calculation of A, we note that the curve $\ell(X_N^x)$ can be built up stage by stage by the recurrence relation (9.6). Suppose that the recurrence has gone as far as $N = n$, and we look for the expected increase in area, say q_n, in going from $N = n$ to $N = n + 1$. We shall have

$$A = \sum_{n=0}^{N-1} q_n. \tag{12.4}$$

The discontinuities of $\ell(X_n^x)$ occur at

$$y_{1,n} < y_{2,n} < \cdots < y_{I(n),n} \tag{12.5}$$

in accordance with (9.8); and the recurrence arises from adding a new point x_{n+1} uniformly distributed over $[0, 1]$. The added area will be $(y_{j+1} - x_{n+1})$, where y_{j+1} is the first term in (12.5) which exceeds x_{n+1}. Given that x_{n+1} falls in (y_j, y_{j+1}) the conditional expected additional area is $\frac{1}{2}(y_{j+1} - y_j)$; and the probability that x_{n+1} does fall in (y_j, y_{j+1}) is $(y_{j+1} - y_j)$. Hence

$$q_n = \sum_{j=0}^{I(n)} \tfrac{1}{2}(y_{j+1} - y_j)^2, \tag{12.6}$$

where we have written y_i for $y_{i,n}$ and taken $y_0 = 0$, $y_{I(n)+1} = 1$.

Now consider the contribution to q_n from an interval $(x, x + dx)$. Here dx is small; but we suppose that n is large enough to make ν large, where ν is the number of y_i in $(x, x + dx)$. It is reasonable to assume that the distribution of the discontinuities y_i is locally homogeneous and random: I shall refer to this as "assumption α," and will return later to discuss it. Thus, in $(x, x + dx)$ the y_i behave as though they were independently and uniformly distributed over this interval of length dx. However, it is known [19] that when ν points are uniformly and independently distributed over a unit interval, thus dividing it into $\nu + 1$ subintervals, the sum of squares of the lengths of these subintervals has an expected value $2/(\nu + 2)$. Multiplying by $(dx)^2$ to allow for the scale factor between $(x, x + dx)$ and a unit interval, writing $2/\nu$ instead of $2/(\nu + 2)$ because ν is large, and incorporating the factor $\frac{1}{2}$ in (12.6) we see that the contribution to q_n from $(x, x + dx)$ is

$$(dx)^2/\nu. \tag{12.7}$$

However ν is the number of (unit height) steps of $\ell(X_n^x)$ in $(x, x + dx)$; so

$$\nu = \frac{\partial \ell(X_n^x)}{\partial x}\, dx. \tag{12.8}$$

Inserting this into (12.7) and integrating to collect together all possible intervals $(x, x + dx)$ we get

$$q_n = \int_0^1 dx \bigg/ \frac{\partial \ell(X_n^x)}{\partial x}. \tag{12.9}$$

Thus (12.1) and (12.9) yield

$$q_n = \int_0^1 \frac{dx}{(c\sqrt{n})(\frac{1}{2}x^{-1/2})} = \frac{4}{3c\sqrt{n}}. \tag{12.10}$$

By (12.4) we get

$$A \sim \sum_{n=n_0}^{N-1} \frac{4}{3c\sqrt{n}} = \frac{8\sqrt{N}}{3c} + o(\sqrt{N}). \tag{12.11}$$

Here the summation has to begin at some large value $n = n_0$, because we have assumed ν large. Equating (12.3) and (12.11) we find that

$$c = 2. \tag{12.12}$$

The mathematical novice will doubtless be appalled at this argument: he will complain that in (12.8) I have differentiated a step function, and in (12.9) I have integrated the reciprocal of this derivative; and that, if the reciprocal of the derivative of a step function has any meaning, it must be infinite everywhere, except for zero values at the positions of the steps. And he will also point to the looseness which sometimes envisages random values and sometimes their expected values and which slips carelessly from one to the other. Experienced mathematicians will feel no such qualms. With one exception, all the steps in the argument can be made rigorous. To sketch very briefly the necessary amendments we first choose an arbitrary $\varepsilon > 0$ and then find a partition of $[0, 1]$ into equal intervals of length $\delta = \delta(\varepsilon)$ such that the Darboux sum

$$\sum_{j=1}^{1/\delta} (j\delta)^{1/2}\delta \tag{12.13}$$

approximates

$$\int_0^1 x^{1/2}\, dx \tag{12.14}$$

to within ε. The intervals $(j\delta, j\delta + \delta)$ can then be used in place of $(x, x + dx)$. The derivative in (12.8) is replaced by a finite difference, which will behave itself if $n \geqq n_0(\varepsilon)$. And N in (12.11) can be chosen large in comparison with n. Eventually we shall find that c differs from 2 by a fixed multiple of ε; and (12.12) will result from the arbitrariness of ε.

The only gap in the argument which I cannot yet fill in rigorously is a justification of assumption α. Unfortunately this assumption is crucial to the argument; for, if we replace the 2 in the numerator of $2/(\nu + 2)$ by some other constant, the ultimate value of c will be altered correspondingly. Of course, Cauchy's inequality shows that the sum of squares in question cannot be less than $1/(\nu + 1)$: this leads to a rigorous proof that $c \geqq \sqrt{2}$, a result which is not as good as $c \geqq \frac{1}{2}\pi$ obtained in Theorem 7. Assumption α is actually a stronger assumption than one needs to arrive at $2/(\nu + 2)$; and, at one stage in the development, I thought that it might be possible to justify an adequate weaker assumption by entering Laplace transformed space at an appropriate moment and utilizing the results of Section 13. (This would have involved an appropriate transformation of the x-axis, and a more complicated integral instead of (12.4); but that would only call for a few technical adjustments of a fairly simple kind.) However, the manipulation in Laplace transformed space (see the end of Section 16) has proved more slippery and difficult than I first thought; and so far I cannot provide a rigorous proof of (12.12). However, I should be very surprised if (12.12) is false.

There is a rather treacherous variant of this method—treacherous because, unlike the foregoing, it hides an elusive mixture of conditional expectations that are not at all easily amenable to rigor. The conditional expectation of $\ell(X_{N+1})$ given $\ell(X_N^x)$ is $1 - y_{I(N),N}$, since $\ell(X_{N+1}) - \ell(X_N)$ can only take the values 0 and 1 and the latter value occurs if and only if $y_{I(N)N} \leqq x_{N+1} \leqq 1$. Hence

$$E\ell(X_{N+1}) - E\ell(X_N) = E(1 - y_{I(N),N}). \tag{12.15}$$

The left side of (12.15) is the result of differencing $c\sqrt{N} + o(\sqrt{N})$ with respect to N; and ought to be about $\frac{1}{2}c/\sqrt{N}$ if the error term is smooth (this sort of difficulty ought to be surmountable by reversing the argument and summing the differences over N as in (12.4)). The right side of (12.15) is the displacement in x near $x = 1$ just sufficient to ensure unit increase in $\ell(X_N^x)$, and should be nearly equal to the reciprocal of $\partial\ell(X_N^x)/\partial x$ at $x = 1$, namely $2/c\sqrt{N}$ from (12.1). Thus

$$\tag{12.16} \tfrac{1}{2}c/\sqrt{N} = 2/c\sqrt{N},$$

which once again yields (12.12).

This argument can be made a little bit more plausible by considering what happens at the discontinuity $x = y_{iN}$. By the same argument as leads to (12.15), we have

$$\tag{12.17} E\ell(X_{N+1}^x) - \ell(X_N^x) = y_{i,N} - y_{i-1,N} \qquad (x = y_{i,N}),$$

where the expectation in (12.17) is conditional on $\ell(X_N^x)$ being given. Now $\ell(X_N^x)$ regarded as a function of x has a jump of height 1 at $x = y_{i,N}$. Hence summing (12.17) over $i = 1, 2, \cdots, I(N)$, we have the Stieltjes integral

$$\tag{12.18} \int_{x=0}^{1} \{E\ell(X_{N+1}^x) - \ell(X_N^x)\}\, d\ell(X_N^x) = 1.$$

Here we have written 1 for the upper limit of integration and for the right side: strictly it should have been $y_{I(N),N}$; but this is very nearly equal to 1. If we now regard $\ell(X_N^x)$ as a continuous function of N, for example by making it piecewise linear between the original integer values of N, we can write (12.18) as

$$\tag{12.19} E\int_{x=0}^{1} \frac{\partial\ell(X_N^x)}{\partial N}\, d\ell(X_N^x) = 1.$$

Suppose that somehow we can take expected values over X_N^x (and it is not clear how to do so) in such a way that $\ell(X_N^x)$ may be replaced by its asymptotic value $c\sqrt{Nx}$. Then this would give

$$\tag{12.20} \int_{x=0}^{1} \frac{\partial}{\partial N}(c\sqrt{Nx})\frac{\partial}{\partial x}(c\sqrt{Nx})\, dx = 1;$$

whereupon an easy computation leads from (12.20) to (12.12).

In Section 1 I said that published proofs differ from the arguments of their gestation. Sections 10, 11, and 12 may illustrate the kind of preliminary thinking from which a finished proof might be derived. Each appears in a different stage of development and none has reached fruition. As a matter of fact I have presented them in reverse chronology. Section 12, apart from the much younger final variant, is about six weeks old and comes nearest to being the framework for a rigorous argument. Section 11 is about two weeks old, and Section 10 is only two days old. These ages are reflected in the relative coherence or incoherence of the text. If any one of them had reached the stage of a watertight argument, the others would have been discarded. Moreover the necessary epsilontics of a proof would have shrouded the underlying ideas and their origins.

13. The joint characteristic function

It is always worth trying to solve a problem by brute force. If the problem is a difficult one, brute force will have slender hopes of success; but, should it succeed, it is likely to yield far more detailed results than general theory ever does. Charles Darwin was fond of saying that scientists should perform damn-fool experiments from time to time: these usually fail, but are a triumph when they come off. He illustrated this precept once by playing the trombone at his tulips, with negative results.

We should be able to deduce almost anything we wanted to know about Ulam's problem if we could obtain a tractable expression for the joint characteristic function of the discontinuities in the Monte Carlo truncated function (9.6). The distribution of ℓ_N is nonparametric in the sense that it does not depend upon the common distribution of the x_i in X; and the distribution of ℓ_N is the same as that of ℓ'_N. There are technical simplifications in supposing that the x_i come from the common distribution

$$P(x_i \leqq x) = 1 - e^{-x}, \qquad 0 \leqq x \leqq \infty, \tag{13.1}$$

and in considering ℓ'_N instead of ℓ_N. The necessary changes to (9.6) and (9.8) are as follows. We define $\bar{X}_N^x$ to be the subsequence of X_N which contains only those $x_i \geqq x$. Thus $\bar{X}_N^0 = X_N$ in particular. We write $\ell'(\bar{X}_N^x)$ for the length of a longest descending subsequence in $\bar{X}_N^x$. Thus (9.6) becomes

$$\ell'(\bar{X}_{N+1}^x) = \begin{cases} \ell'(\bar{X}_N^x), & x_{N+1} < x \leqq \infty \\ \max\{\ell'(\bar{X}_N^x),\ 1 + \ell'(\bar{X}_N^{x_{N+1}})\}, & 0 \leqq x \leqq x_{N+1}. \end{cases} \tag{13.2}$$

Suppose that the steps of $\ell'(\bar{X}_N^x)$ occur at

$$y_{1,N} > y_{2,N} > \cdots > y_{I(N),N}. \tag{13.3}$$

For convenience, we drop the second suffix and extend the range of the first suffix: so (13.3) becomes

$$y_1 \geqq y_2 \geqq \cdots \geqq y_i \geqq \cdots \geqq 0. \tag{13.4}$$

In (13.4) all $y_i = 0$ when $i > I(N)$; and, with probability 1, strict inequality holds for $y_i > y_{i+1}$ when $i \leqq I(N)$. The advantage of (13.4) lies in not having to bother about the value of $I(N)$ in the calculation, since the notation automatically takes care of it. Our aim is to find an expression for

$$\phi_N(\mathrm{s}) = \phi_N(s_1, s_2, \cdots) = E \exp\left\{-\sum_{i=1}^{\infty} s_i y_i\right\} \tag{13.5}$$

where the suffix N in (13.5) recalls the suppressed second suffix in the y_i. We note that the extended y_i, namely those with $i > I(N)$, being zero do not affect the value of $\phi_N(\mathrm{s})$. In (13.5), we would have a joint characteristic function if the s_i were all pure imaginary quantities: it is, however, more convenient to take the s_i to be real and nonnegative; and this will not affect the usefulness of $\phi_N(\mathrm{s})$. We now seek a recurrence relation between ϕ_N and ϕ_{N+1}.

To pass from (13.4), which relates to N, to the corresponding sequence relating to $N + 1$, we have to draw x_{N+1} from the distribution (13.1), and then insert x_{N+1} in place of y_j where j is the smallest integer such that $y_j > x_{N+1}$. If no such j exists, we simply augment the y by x_{N+1} as it stands. This will provide a new sequence

$$y'_1 \geqq y'_2 \geqq \cdots \geqq 0 \tag{13.6}$$

in place of (13.4); from which

$$\phi_{N+1}(\mathbf{s}) = E \exp\left\{ - \sum_{i=1}^{\infty} s_i y'_i \right\}. \tag{13.7}$$

We shall, however, calculate (13.7) in two stages. In the first stage we calculate

$$E^* \exp\left\{ - \sum_{i=1}^{\infty} s_i y'_i \right\}, \tag{13.8}$$

where E^* denotes the conditional expectation given the sequence (13.4). The second stage will complete the process by taking expectations over the sequence (13.4). For typographical convenience we write x in place of x_{N+1}. Thus drawing x from the distribution (13.1), we have

$$\begin{aligned} E^* \exp&\left\{ - \sum_{i=1}^{\infty} s_i y'_i \right\} \\ &= \int_{y_1}^{\infty} e^{-x} \exp\left\{ - s_1 x - \sum_{i=2}^{\infty} s_i y_i \right\} dx \\ &\quad + \sum_{j=2}^{\infty} \int_{y_j}^{y_{j-1}} e^{-x} \exp\left\{ - \sum_{i=1}^{j-1} s_i y_i - s_j x - \sum_{i=j+1}^{\infty} s_i y_i \right\} dx \\ &= \frac{1}{s_1 + 1} \exp\left\{ - (s_1 + 1) y_1 - \sum_{i=2}^{\infty} s_i y_i \right\} \\ &\quad + \sum_{j=2}^{\infty} \frac{1}{s_j + 1} \left[\exp\left\{ - \sum_{i=1}^{j-1} s_i y_i - (s_j + 1) y_j - \sum_{i=j+1} s_i y_i \right\} \right. \\ &\qquad \left. - \exp\left\{ - \sum_{i=1}^{j-1} s_i y_i - (s_j + 1) y_{j-1} - \sum_{i=j+1}^{\infty} s_i y_i \right\} \right]. \end{aligned} \tag{13.9}$$

Now we effect the second stage of the calculation. Using the definition (13.5) with appropriately adjusted values of s, we find

$$\begin{aligned} \phi_{N+1}(\mathbf{s}) &= \frac{1}{s_1 + 1} \phi_N(s_1 + 1, s_2, s_3, \cdots) \\ &\quad + \sum_{j=2}^{\infty} \frac{1}{s_j + 1} [\phi_N(s_1, s_2, \cdots, s_{j-1}, s_j + 1, s_{j+1}, \cdots) \\ &\quad - \phi_N(s_1, s_2, \cdots, s_{j-2}, s_{j-1} + s_j + 1, 0, s_{j+1}, \cdots)]. \end{aligned} \tag{13.10}$$

This recurrence relation can be started by considering the case $N = 0$, in which $y_1 = y_2 = \cdots = 0$. Thus (13.10) holds for $N = 0, 1, 2, \cdots$ and starts from

$$\phi_0(\mathbf{s}) = 1. \tag{13.11}$$

The next problem is to solve the functional recurrence relation (13.10), and a pretty unpleasant relation it looks. Functional equations in one variable can be troublesome, those in infinitely many variables are worse. The only hope, I felt, was to work out the first few values $\phi_0, \phi_1, \phi_2, \cdots$ from (13.10), to guess the general result, and then prove it by induction. This is a good example of brute force mathematics, and it needs plenty of courage and a lot of tedious elementary algebra. It is clear that ϕ_N will be a rational function of $s_1, s_2, \cdots, s_N$. The difficulty is that a rational function of several variables can be written in an enormous number of different ways, and one has to hit on just the right way of writing it before one has much chance of guessing the general result. If the reader does not believe this, let him try solving (13.10) for himself.

I spent two or three days over this job. The expressions ϕ_1 and ϕ_2 are quite easily found; but ϕ_3 was a lot messier, and I had to write it down in many different algebraic forms before I got it into a shape which looked like a reasonable extension of ϕ_1 and ϕ_2. I then guessed what ϕ_N should be like on the basis of ϕ_1, ϕ_2, and ϕ_3, and was able to confirm this guess by induction. This confirmation is quite easy and runs as follows.

Define

$$S_1 = s_1, \quad S_2 = s_1 + s_2, \quad S_3 = s_1 + s_2 + s_3, \cdots \tag{13.12}$$

and

$$S(\mathbf{a}) = \prod_{i=1}^{k} \frac{(S_i + a_{i-1})!}{(S_i + a_i)!}, \tag{13.13}$$

where $\mathbf{a} = (a_0, a_1, \cdots, a_k)$ is a sequence of integers satisfying

$$0 = a_0 < a_1 < a_2 < \cdots < a_k. \tag{13.14}$$

Introduce the functional operators $J_1, J_2, \cdots$ by the definitions

$$J_1 f(s_1, s_2, \cdots) = (s_1 + 1)^{-1} f(s_1 + 1, s_2, s_3, \cdots), \tag{13.15}$$

$$\begin{aligned} J_j f(s_1, s_2, \cdots) = (s_j + 1)^{-1}[&f(s_1, \cdots, s_{j-1}, s_j + 1, s_{j+1}, \cdots) \\ &- f(s_1, \cdots, s_{j-2}, s_{j-1} + s_j + 1, 0, s_{j+1}, \cdots)], \quad j \geqq 2. \end{aligned} \tag{13.16}$$

From (13.10), we have

$$\phi_{N+1}(\mathbf{s}) = \sum_{j=1}^{\infty} J_j \phi_N(\mathbf{s}). \tag{13.17}$$

Now, from (13.13),

$$J_1 S(a_0, a_1, \cdots, a_k) = S(a_0, a_1 + 1, a_2 + 1, \cdots, a_k + 1), \tag{13.18}$$

and

(13.19)
$$J_j S(a_0, a_1, \cdots, a_k) = 0, \qquad j \geqq k + 2.$$

To deal with the remaining cases $2 \leqq j \leqq k + 1$ we adopt the convention that an empty product, like $\Pi_{j=1}^{0}$, is interpreted as 1. We have, for $2 \leqq j \leqq k + 1$,

(13.20) $J_j S(a_0, a_1, \cdots, a_k)$
$$= (s_j + 1)^{-1}\left[\left\{\prod_{i=1}^{j-1} \frac{(S_i + a_{i-1})!}{(S_i + a_i)!}\right\}\left\{\prod_{i=j}^{k} \frac{(S_i + a_{i-1} + 1)!}{(S_i + a_i + 1)!}\right\}\right.$$
$$\left. - \left\{\prod_{i=1}^{j-2} \frac{(S_i + a_{i-1})!}{(S_i + a_i)!}\right\}\left\{\frac{(S_j + a_{j-2} + 1)!}{(S_j + a_{j-1} + 1)!}\right\}\left\{\prod_{i=j}^{k} \frac{(S_i + a_{i-1} + 1)!}{(S_i + a_i + 1)!}\right\}\right]$$
$$= \left\{\prod_{i=1}^{j-2} \frac{(S_i + a_{i-1})!}{(S_i + a_i)!}\right\}\left\{\prod_{i=j}^{k} \frac{(S_i + a_{i-1} + 1)!}{(S_i + a_i + 1)!}\right\}.$$
$$\frac{1}{S_j - S_{j-1} + 1}\left\{\frac{(S_{j-1} + a_{j-2})!}{(S_j + a_{j-1})!} - \frac{(S_j + a_{j-2} + 1)!}{(S_j + a_{j-1} + 1)!}\right\}$$
$$= \left\{\prod_{i=1}^{j-2} \frac{(S_i + a_{i-1})!}{(S_i + a_i)!}\right\}\left\{\prod_{i=j}^{k} \frac{(S_i + a_{i-1} + 1)!}{(S_i + a_i + 1)!}\right\}.$$
$$\sum_{r=a_{j-2}+1}^{a_{j-1}} \frac{(S_{j-1} + a_{j-2})!}{(S_{j-1} + r)!} \frac{(S_j + r)!}{(S_j + a_{j-1} + 1)!}$$
$$= \sum_{r=a_{j-2}+1}^{a_{j-1}} S(a_0, a_1, \cdots, a_{j-2}, r, a_j + 1, a_{j+1} + 1, \cdots, a_k + 1).$$

Since

(13.21)
$$\phi_1(\mathrm{s}) = (s_1 + 1)^{-1} = S(0, 1)$$

we deduce from (13.17), (13.18), (13.9) and (13.20) that

(13.22)
$$\phi_N(\mathrm{s}) = \sum_{\mathbf{a}}^{N} c(\mathbf{a}) S(\mathbf{a}),$$

where $c(\mathbf{a})$ is a positive integer depending on the sequence $\mathbf{a}$, and $\Sigma_{\mathbf{a}}^{N}$ denotes summation over all integer sequences $\mathbf{a}$ satisfying

(13.23)
$$0 = a_0 < a_1 < a_2 < \cdots < a_k = N.$$

Here k may have any integer value provided $1 \leqq k \leqq N$. Thus there are 2^{N-1} sequences satisfying (13.23); and accordingly $\phi_N(\mathrm{s})$ is a linear combination of 2^{N-1} functions like (13.13), the coefficients of the linear combination being positive integers. This exhibits the functional form of ϕ_N; and to make further progress we need to study the coefficients $c(\mathbf{a})$.

14. Recurrence relations for the coefficients c (a)

The operators $J_1, J_2, \cdots$ are linear operators on the positive integral orthant of the vector space spanned by the functions (13.13). The vectors **a** can be taken as a natural representation of the basis of the orthant. In deriving recurrence relations for the coefficients $c(\mathbf{a})$, it suffices to study the mapping induced on the representation by the linear operator

$$J = \sum_{j=1}^{\infty} J_j. \tag{14.1}$$

Let the set of vectors **a** which are sequences of integers satisfying

$$0 = a_0 < a_1 < \cdots < a_k = N \tag{14.2}$$

be denoted by A_N. Given a vector $\mathbf{a} \in A_N$ and an integer r satisfying $0 \leqq r \leqq N$, we define a mapping T_r from A_N to A_{N+1}:

$$T_r \mathbf{a} = T_r(a_0, a_1, \cdots, a_k) = \mathbf{b} = (b_0, b_1, \cdots, b_{k'}), \tag{14.3}$$

where the b_i are obtained as follows. Determine the smallest integer j such that $r \leqq a_j$. Then (ignoring any empty instruction like $0 \leqq i < 0$) put

$$\begin{aligned} b_i &= a_i, & 0 \leqq i < j \\ b_j &= r, & \\ b_i &= a_i + 1, & j < i \leqq k. \end{aligned} \tag{14.4}$$

If this process results in $b_k = N + 1$, the process is complete; if it results in $b_k < N + 1$, put $k' = k + 1$ and $b_{k'} = N + 1$. A vector $\mathbf{a} \in A_N$ is an inverse image of $\mathbf{b} \in A_{N+1}$ if there exists an r, $0 \leqq r \leqq N$ such that $T_r\, \mathbf{a} = \mathbf{b}$. We write $T^{-1}(\mathbf{b})$ for the set of all inverse images of **b**. Comparison of (14.1), (14.3), and (14.4) with (13.18) and (13.20) now shows that

$$c(\mathbf{b}) = \sum_{\mathbf{a} \in T^{-1}(\mathbf{b})} c(\mathbf{a}). \tag{14.5}$$

This is the desired recurrence relation for the coefficients. The recurrence starts from

$$c(0, 1) = 1, \tag{14.6}$$

in accordance with (13.21).

The rules by which the recurrence (14.5) operates are somewhat complicated and a numerical example will help to clarify them. We use the rules to construct ϕ_2 from ϕ_1, ϕ_3 from ϕ_2, and so on. Suppose that we have got as far as ϕ_4, and we now wish to calculate ϕ_5 from ϕ_4. To simplify the notation we shall write typically

$$S_{034} = S(0, 3, 4). \tag{14.7}$$

Here the inital zero suffix serves to distinguish $S_{01} = S(0, 1)$, for example, from the variable S_1. In this notation ϕ_4 is given by

$$\phi_4(\mathrm{s}) = S_{04} + 5S_{014} + 5S_{024} + 3S_{034} + 3S_{0124} + 3S_{0134} + 3S_{0234} + S_{01234}. \tag{14.8}$$

The calculation proceeds by means of Table IV. The first column contains the

TABLE IV

CALCULATION OF ϕ_5 FROM ϕ_4

a	$c(\mathbf{a})$	$T_0\mathbf{a}$	$T_1\mathbf{a}$	$T_2\mathbf{a}$	$T_3\mathbf{a}$	$T_4\mathbf{a}$
04	1	05	015	025	035	045
014	5	025	015	0125	0135	0145
024	5	035	015	025	0235	0245
034	3	045	015	025	035	0345
0124	3	0235	0135	0125	01235	01245
0134	3	0245	0145	0125	0135	01345
0234	3	0345	0145	0245	0235	02345
01234	1	02345	01345	01245	01235	012345

suffices $\mathbf{a} \in A_4$ appearing in ϕ_4, and the second column gives the corresponding coefficients $c(\mathbf{a})$ taken from (14.8). The last five columns tabulate $T_r\,\mathbf{a}$, for $r = 0, 1, \cdots, 4$. In general when calculating ϕ_{N+1} from ϕ_N the table will contain 2^{N-1} rows and $N + 3$ columns for $\mathbf{a}$, $c(\mathbf{a})$, and $T_r\,\mathbf{a}(r = 0, 1, \cdots, N)$. To calculate $T_2(034)$, for example, we note that 3 is the least integer in $\mathbf{a} = (034)$ which exceeds $r = 2$. The integer 3 is accordingly reduced to 2; all preceding integers are unaltered; and all succeeding integers are increased by 1. Thus $T_2(034) = (025)$. The exception to this rule arises when it would lead to a vector which did not end in $N + 1 = 5$. In that event $N + 1$ is added to the end of the vector as a final coordinate. Thus $T_2(014) = 0125$, and not 012 under the unamended rule. Examination of the various entries in Table IV should make everything clear. The entries in the body of the table are the vectors $\mathbf{b}$ in A_{N+1}; and to calculate $c(\mathbf{b})$ we add together the $c(\mathbf{a})$ entries, in the second column, for each row in which $\mathbf{b}$ occurs. For example, from rows 2, 5, and 6,

$$c(0135) = 5 + 3 + 3 = 11. \tag{14.9}$$

This leads to

$$\begin{aligned}\phi_5(\mathrm{s}) = S_{05} &+ 14S_{015} + 14S_{025} + 9S_{035} + 4S_{045} + 11S_{0125} \\ &+ 11S_{0135} + 11S_{0145} + 11S_{0235} + 11S_{0245} \\ &+ 6S_{0345} + 4S_{01235} + 4S_{01245} + 4S_{01345} \\ &+ 4S_{02345} + S_{012345}.\end{aligned} \tag{14.10}$$

In this way I calculated $\phi_1, \phi_2, \cdots, \phi_7$; and Mr. A. Izenman checked my calculations and extended them to ϕ_8 and ϕ_9. The size of the calculation doubles for each new ϕ; and ϕ_9 is about as far as one can go with paper and pencil.

Things could go further with a computer; but even a large computer would feel its resources strained round about ϕ_{20}; and the method is clearly impractical for $N \geqq 30$.

What is needed is some algebraic apparatus, say generating functions, to carry the work to larger values of N. I have not yet succeeded in constructing such apparatus. However, examination of the first few values of ϕ_N led to the formulation of the functional form of ϕ, namely the functions (13.13); and, in the same spirit, we can look at the numerical properties of the first few coefficients $c(\mathbf{a})$ in the hope of spotting some general pattern. In combinatorial work especially, but also in other areas of mathematics, I find it very helpful to study the numerical properties of particular cases. If the numerical data are extensive, one must summarize in some way that will fruitfully reveal the intrinsic pattern. Research experience, rather than undergraduate learning, seems to be the only road to cultivating an instinct for the fruitful choice of good summarizing quantities.

15. Properties of the coefficients c (a)

Even for small values of N, there is an unwieldy amount of data associated with the coefficients $c(\mathbf{a})$; and some method of summarizing it is advisable. As an ad hoc device, guided by a mixture of instinct and experience, I decided to look at the quantities $\alpha_{p,q}^{(N)}$ defined by

$$\alpha_{p,q}^{(N)} = \sum_{a_p = q,\, a_k = N} c(a_0, a_1, \cdots, a_k), \qquad 1 \leqq p \leqq q \leqq N. \tag{15.1}$$

For example, from (14.10),

$$\alpha_{24}^{(5)} = 11 + 11 + 6 = 28. \tag{15.2}$$

The results are tabulated in Tables V to IX, with row and column totals, for $N = 1, 2, \cdots, 7$.

TABLE V

VALUE OF $\alpha_{p,q}^{(1)}$

	$q = 1$	
$p = 1$	1	1
	1	

TABLE VI

VALUES OF $\alpha_{p,q}^{(2)}$

	$q = 1$	2	
$p = 1$	1	1	2
2		1	1
	1	2	

TABLE VII

VALUES OF $\alpha_{p,q}^{(3)}$

	$q = 1$	2	3	
$p = 1$	3	2	1	6
2		1	4	5
3			1	1
	3	3	6	

TABLE VIII

VALUES OF $\alpha_{p,q}^{(4)}$

	$q = 1$	2	3	4	
$p = 1$	12	8	3	1	24
2		4	6	13	23
3			1	9	10
4				1	1
	12	12	10	24	

TABLE IX

VALUES OF $\alpha_{p,q}^{(5)}$

	$q = 1$	2	3	4	5	
$p = 1$	60	40	15	4	1	120
2		20	30	28	41	119
3			5	12	61	78
4				1	16	17
5					1	1
	60	60	50	45	120	

Have you spotted any numerical patterns yet? If you wish to test your skill at pattern spotting, *do not turn over the page* until you have first had a very good look at tables V to IX inclusive and formed your own conjectures about the corresponding numerical patterns for the next two cases ($N = 6$ and $N = 7$). These next two cases are covered by Tables X and XI on the next page.

Scrutiny of these tables suggests certain interesting patterns. In the first place it appears to be true that

(15.3) $$\alpha_{p,q}^{(N)} = N\alpha_{p,q}^{(N-1)}, \qquad 1 \leqq p \leqq q \leqq N - 2.$$

On the other hand, (15.3) is certainly *not* true for $q = N - 1$ or $q = N$. Secondly, if we denote the row totals by

(15.4) $$\alpha_p^{(N)} = \sum_{q=p}^{N} \alpha_{p,q}^{(N)},$$

then it appears to be true that

(15.5) $$\alpha_1^{(N)} = N!$$

TABLE X

VALUES OF $\alpha^{(6)}_{p,q}$

	$q = 1$	2	3	4	5	6	
$p = 1$	360	240	90	24	5	1	720
2		120	180	168	120	131	719
3			30	72	105	381	588
4				6	20	181	207
5					1	25	26
6						1	1
	360	360	300	270	251	720	

TABLE XI

VALUES OF $\alpha^{(7)}_{p,q}$

	$q = 1$	2	3	4	5	6	7	
$p = 1$	2520	1680	630	168	35	6	1	5040
2		840	1260	1176	840	495	428	5039
3			210	504	735	830	2332	4611
4				42	140	276	1821	2279
5					7	30	421	458
6						1	36	37
7							1	1
	2520	2520	2100	1890	1757	1638	5040	

and

(15.6) $$\alpha^{(N)}_p = \alpha^{(N)}_{p-1} - \alpha^{(N)}_{p-1,N}, \qquad 2 \leqq p \leqq N.$$

We shall prove later that (15.3), (15.5), and (15.6) are indeed true in general. For the moment we only note that they were originally obtained on the empirical evidence of Tables V, VI, $\cdots$, XI; and that they have the following important implication: all the numbers $\alpha^{(N)}_{p,q}$ can be reconstructed from a knowledge of the last columns $\alpha^{(N)}_{p,N}$ only. For suppose that we are given $\alpha^{(n)}_{p,n}$ for all p, n satisfying $1 \leqq p \leqq n \leqq N$, and suppose that we have so far managed from these to reconstruct $\alpha^{(n)}_{pq}$ for all p, q, n satisfying $1 \leqq p \leqq q \leqq n \leqq N - 1$. Then we can reconstruct the first $N - 2$ columns of the table $\alpha^{(N)}_{p,q}$ by use of (15.3). The Nth column of the table has been given us; and we can calculate the row totals of the table by successive use of (15.5) and (15.6). We can then fill in the $(N - 1)$th column, since it is the only missing column and we know the row totals. The assertion about reconstruction now follows by induction upon N.

Thus it is enough to study the quantities

(15.7) $$\beta^{(N)}_p = \alpha^{(N)}_{p,N},$$

which we now tabulate for $1 \leqq p \leqq N \leqq 9$.

I have said that Table XII is sufficient for the reconstruction of the earlier tables; but it is much more important than this, and actually it contains the complete solution of Ulam's problem (or rather, we should possess the complete

TABLE XII

VALUES OF $\beta_p^{(N)}$

	N = 1	2	3	4	5	6	7	8	9
p = 1	1	1	1	1	1	1	1	1	1
2		1	4	13	41	131	428	1429	4861
3			1	9	61	381	2332	14337	89866
4				1	16	181	1821	17557	167080
5					1	25	421	6105	83029
6						1	36	841	16465
7							1	49	1513
8								1	64
9									1
	1	2	6	24	120	720	5040	40320	362880

solution if we knew the complete form of Table XII instead of its first 9 columns only). At first I did not realize the significance of Table XII: it merely evolved as a study of certain numerical patterns associated with the coefficient $c(\mathbf{a})$. It was not until I had calculated Tables V, VI, $\cdots$, IX and hence the first 5 columns of Table XII that I understood what Table XII meant. When its meaning dawned on me, I decided to calculate Tables X and XI, and Mr. Izenman extended this to $N = 8$ and $N = 9$, and at that stage Table XII emerged in its present form. I shall explain the meaning of Table XII in the next section: for the moment, let us look at the numerical patterns in Table XII.

Evidently we have for the first row

(15.8) $$\beta_1^{(N)} = 1, \qquad N \geqq 1.$$

The second row is not so simple; but it turns out that

(15.9) $$\beta_2^{(N)} = \frac{(2N)!}{N!(N+1)!} - 1, \qquad N \geqq 2,$$

and this can be proved in general. There is also a fairly clear pattern in diagonals near the bottom of the table:

(15.10) $$\beta_N^{(N)} = 1, \qquad N \geqq 1$$

(15.11) $$\beta_{N-1}^{(N)} = (N-1)^2, \qquad N \geqq 2.$$

Mr. Izenman discovered the formulae for the next two diagonals,

(15.12) $$\beta_{N-2}^{(N)} = \tfrac{1}{2}N(N-1)(N-2)(N-3) + 1, \qquad N \geqq 3$$

and

(15.13) $$\begin{aligned}\beta_{N-3}^{(N)} = 1 + \tfrac{1}{3}N(N-1)[&-115 + 57N - 10N(N-1) \\ &+ \frac{7}{4!}N(N-1)(N-2)(N-3) \\ &+ \frac{1}{5!}N(N-1)(N-2)(N-3)(N-4)], \qquad N \geqq 4.\end{aligned}$$

Equations (15.10), (15.11), and (15.12) can be proved in general; but (15.13) has not been proved in general and may not have been written in the most transparent form. This question of transparency is elusive. For example, when one knows the reason why (15.12) is true, it is appropriate to write (15.12) in the form

$$(15.14)\qquad \beta_{N-2}^{(N)} = \binom{N-1}{2} + \left[\binom{N-1}{2} - 1\right]\left[2\binom{N-1}{2} - 1\right], \qquad N \geqq 3;$$

and it seems to be more or less an algebraic accident that the right side of (15.14) happens to simplify to the right side of (15.12). If (15.13) really is as complicated as it looks, then an attempt to find exact general formulae for the quantities $\beta_p^{(N)}$ would seem out of the question, and we might have to be content with approximations or asymptotic formulae. This issue remains unsettled.

16. Interpretation of the coefficients c (a)

The following interpretation of the coefficients $c(\mathbf{a})$ emerged gradually from an attempt to prove (15.3), which at that early stage was merely a conjecture based on the numerical evidence of Tables V, VI, $\cdots$, IX. In deriving the formulae for ϕ_N we used descending subsequences of $\{x_1, x_2, \cdots\}$, where the x_i came from the exponential distribution (13.1). This, however, was merely a device for easing the analysis and obtaining manageable functions (13.13). However, the coefficients $c(\mathbf{a})$ are much more deeply implicated in the combinatorial structure of the problem; and, to interpret them, we return to the original formulation of the problem, namely ascending subsequences of random permutations of the integers $\{1, 2, \cdots, N\}$.

We set up a method of coding these permutations. Suppose that $\pi = \{\pi_1, \pi_2, \cdots, \pi_N\}$ is a given permutation of $\{1, 2, \cdots, N\}$. Define

$$(16.1)\qquad a_i = a_i(\pi), \qquad i = 1, 2, \cdots, \ell$$

to be the greatest integer j such that $\{\pi_1, \pi_2, \cdots, \pi_j\}$ has a longest ascending subsequence of length i. In the definition (16.1), $i = 1, 2, \cdots, \ell$ where ℓ is the length of a longest ascending subsequence in $\pi = \{\pi_1, \pi_2, \cdots, \pi_N\}$. Also define $a_0 = 0$; and write

$$(16.2)\qquad \mathbf{a}(\pi) = \big(a_0, a_1(\pi), \cdots, a_\ell(\pi)\big).$$

For example, if $N = 9$ and

$$(16.3)\qquad \pi = \{8\ 9\ 1\ 4\ 3\ 6\ 5\ 7\ 2\},$$

then

$$(16.4)\qquad \mathbf{a}(\pi) = (0\ 1\ 5\ 7\ 9).$$

It is evident that the final coordinate in $\mathbf{a}(\pi)$ must always equal N, the number of elements in π. For a given vector $\mathbf{a}$, let $\gamma(\mathbf{a})$ denote the number of permutations π such that $\mathbf{a}(\pi) = \mathbf{a}$. We are going to prove that

$$\gamma(\mathbf{a}) = c(\mathbf{a}). \tag{16.5}$$

If $N = 1$, the only possible permutation is $\pi = \{1\}$ and $a(\pi) = (0, 1)$. Hence $\gamma(0, 1) = 1 = c(0, 1)$ and (16.5) is true for $N = 1$. Now assume that (16.5) is true for N. Consider a given permutation $\pi = \{\pi_1, \pi_2, \cdots, \pi_N\}$. This can be converted to a permutation of $\{1, 2, \cdots, N + 1\}$ by inserting $N + 1$ in $N + 1$ available places. We write these extended permutations as

$$\begin{aligned} T_0\pi &= \{N + 1, \pi_1, \pi_2, \cdots, \pi_N\} \\ T_1\pi &= \{\pi_1, N + 1, \pi_2, \cdots, \pi_N\} \\ T_2\pi &= \{\pi_1, \pi_2, N + 1, \cdots, \pi_N\} \\ &\text{-----------------} \\ T_N\pi &= \{\pi_1, \pi_2, \cdots, \pi_N, N + 1\}. \end{aligned} \tag{16.6}$$

However, by (14.3) and (14.4), we have

$$\mathbf{a}(T_r\,\pi) = T_r\,\mathbf{a}(\pi), \qquad r = 0, 1, \cdots, N. \tag{16.7}$$

This holds for all π; and hence (16.5) is true for $N + 1$ in place of N. Thus (16.5) is generally true by induction on N.

But now (15.1) and (15.7), taken together with (16.2), prove that $\beta_p^{(N)}$ is the number of permutations of $\{1, 2, \cdots, N\}$ which contain a longest ascending subsequence of length p. Hence

$$P\{\ell(X_N) = p\} = \beta_p^{(N)}/N!, \tag{16.8}$$

which provides the distribution of the random variable $\ell(X_N)$. This explains the importance of Table XII.

With these preliminaries settled, we can now prove (15.3). Consider a given permutation $\pi = \{\pi_1, \pi_2, \cdots, \pi_N\}$. For $r = 1, 2, \cdots, N + 1$ define

$$U_r\pi = \{\pi_1^*, \pi_2^*, \cdots, \pi_N^*, r\} \tag{16.9}$$

where $\pi_i^* = \pi_i$ or $\pi_i^* = \pi_i + 1$ accordingly as $\pi_i < r$ or $\pi_i \geqq r$. If

$$\mathbf{a}(\pi) = (a_0, a_1, \cdots, a_{\ell-1}, N) \tag{16.10}$$

then

$$\mathbf{a}(U_r\pi) = \begin{cases} (a_0, a_1, \cdots, a_{\ell-1}, N + 1), & r = 1, 2, \cdots, N \\ (a_0, a_1, \cdots, a_{\ell-1}, N, N + 1), & r = N + 1. \end{cases} \tag{16.11}$$

This gives the important identity

$$\begin{aligned} &(N + 1)c(a_0, a_1, \cdots, a_{\ell-1}, N) \\ &\quad = c(a_0, a_1, \cdots, a_{\ell-1}, N + 1) + c(a_0, a_1, \cdots, a_{\ell-1}, N, N + 1), \end{aligned} \tag{16.12}$$

because π is an arbitrary permutation of $\{1, 2, \cdots, N\}$. Since $a_{\ell-1} \leqq N - 1$, we now see that

$$\alpha_{p,q}^{(N+1)} = (N+1)\alpha_{p,q}^{(N)}, \qquad 1 \leqq p \leqq q \leqq N-1, \tag{16.13}$$

which is (15.3) with $N + 1$ in place of N.

A few of the coefficients $c(\mathbf{a})$ can be determined explicitly. For example, consider the permutations which are coded by the vector

$$\mathbf{a} = (0, r, r+1, r+2, \cdots, N), \tag{16.14}$$

that is to say with $a_i = i + r - 1$, for $i = 1, 2, \cdots, N - r + 1$; all have

$$\pi_1 > \pi_2 > \cdots > \pi_r = 1 < \pi_{r+1} < \pi_{r+2} < \cdots < \pi_N, \tag{16.15}$$

and *vice versa*. But we can choose any set of $r - 1$ elements from $\{2, 3, \cdots, N\}$ and arrange them in descending order to give $\pi_1, \pi_2, \cdots, \pi_{r-1}$. Therefore

$$c(0, r, r+1, r+2, \cdots, N) = \binom{N-1}{r-1}. \tag{16.16}$$

From (16.12) and (16.16) we deduce

$$c(0, r, r+1, r+2, \cdots, N-2, N) = \left\{\frac{N^2 - N(r+1) + 1}{r-1}\right\}\binom{N-2}{r-2}. \tag{16.17}$$

For example,

$$c(0345) = \binom{4}{2} = 6, \tag{16.18}$$

and

$$c(035) = \left(\frac{25 - 20 + 1}{2}\right)\binom{3}{1} = 9; \tag{16.19}$$

and these confirm the coefficients of S_{0345} and S_{035} in (14.10).

The identity (16.12) shows that our problem would be solved, at least in principle, if for each N we knew the values of the 2^{N-2} coefficients $c(\mathbf{a})$ in which $a_{\ell-1} = N - 1$, $a_\ell = N$. This suggests that we ought to look for further identities like (16.12) which would successively reduce the problem to a determination of coefficients with

$$\begin{aligned} &a_{\ell-1} = N - 1, a_\ell = N \\ &a_{\ell-2} = N - 2, a_{\ell-1} = N - 1, a_\ell = N \\ &\text{-----------------------} \end{aligned} \tag{16.20}$$

until we reach known coefficients of the form (16.16). However this attractive possibility has so far eluded me. Nor have I made any substantial progress towards an explanation of the spectrum of values assumed by the coefficients

$c(\mathbf{a})$. Table XIII shows the number of distinct values taken by $c(\mathbf{a})$; and this number is noticeably smaller than the number of coefficients.

TABLE XIII

DATA ON THE COEFFICIENT SPECTRUM

Value of N:	1	2	3	4	5	6	7	8	9
Number of coefficients $c(\mathbf{a})$	1	2	4	8	16	32	64	128	256
Number of distinct coefficients $c(\mathbf{a})$	1	1	2	3	6	9	16	29	55

For purposes of reference and to spare other investigators the labor of recalculating the coefficients $c(\mathbf{a})$ for $N \leqq 9$, I give in Table XIV a condensed list of the coefficients for $N = 9$. Coefficients for $N \leqq 8$ can be easily recovered from Table XIV by means of the identity (16.12). To save space the coefficients are simply listed in natural order beginning with $c(01)$ and ending with $c(0123456789)$; and an entry such as u^n means that the coefficient u occurs n times consecutively in this position of the list, and semicolons separate coefficients for different values of ℓ: thus, in this notation, (14.10) would take the compact form 1; 14^2, 9, 4; 11^5, 6; 4^4; 1.

TABLE XIV

LIST OF COEFFICIENTS FOR $N = 9$

1; 1430^2, 1001, 572, 275, 110, 35, 8; 6529^3, 6031, 5035, 4168, 2431, 6529^2, 6031, 5035, 3751, 2431, 3820, 3772, 3322, 2536, 1672, 1609^2, 1321, 913, 520^2, 400, 133^2, 28; 4364^4, 4280, 4028, 4364^3, 4280, 4028, 4364^2, 4280, 4028, 3812^2, 3644, 2876^2, 1475, 4364^3, 4280, 4028, 4364^2, 4280, 4028, 3812^2, 3644, 2876^2, 1892, 2339^3, 2255, 2339^2, 2303, 1871^2, 1271, 866^5, 650, 245^3, 56; 1405^{14}, 1363, 1405^9, 1363, 1405^5, 1363, 1201^3, 877, 1405^9, 1363, 1405^5, 1363, 1201^3, 877, 730^9, 568, 259^4, 70; 314^{34}, 266, 314^{14}, 266, 161^5, 56; 55^{27}, 28; 8^8; 1.

We now return to a further consideration of assumption α in Section 12, where we had to calculate the quantity q_n in (12.6). Actually we shall change the ground a little by supposing that, instead of sampling the x_i in X from the uniform distribution on $[0, 1]$, we are sampling from (13.1) and looking at the distribution of ℓ'_N. The necessary technical adaptation to pass from one form of the problem to the other is simple, merely a suitable transformation of the x-axis. Also (13.4) will hold in place of (12.5) and we write

$$z_i = y_i - y_{i+1}. \qquad (16.21)$$

We want to calculate

$$q_N = \frac{1}{2} \sum_{i=1}^{\infty} z_i^2. \qquad (16.22)$$

From (13.5) and (13.12) we have

$$(16.23)\qquad \phi_N(s) = E\exp\left\{-\sum_{i=1}^{\infty} s_i y_i\right\} = E\exp\left\{-\sum_{i=1}^{\infty} S_i(y_i - y_{i+1})\right\}$$
$$= E\exp\left\{-\sum_{i=1}^{\infty} S_i z_i\right\}.$$

Hence, from (13.13) and (13.22),

$$(16.24)\qquad E\exp\left\{-\sum_{i=1}^{\infty} S_i z_i\right\} = \sum_{\mathbf{a}}^{N} c(\mathbf{a}) \prod_{i=1}^{k} \frac{(S_i + a_{i-1})!}{(S_i + a_i)!}.$$

If the right side of (16.24) had just one term in the sum, this would establish the independence of the z_i required under assumption α. As things stand, it is no more than suggestive.

We can go a little further by differentiating (16.24) twice, and then putting $S_1 = S_2 = \cdots = 0$. Thus

$$(16.25)\qquad q_N = \frac{1}{2}\left[\sum_{i=1}^{\infty} \frac{\partial^2}{\partial S_i^2} E\exp\left\{-\sum_{j=1}^{\infty} S_j z_j\right\}\right]_{\mathbf{s}=\mathbf{0}}$$
$$= \frac{1}{N!}\sum_{\mathbf{a}}^{N} c(\mathbf{a}) v(\mathbf{a}),$$

where

$$(16.26)\qquad v(a_0, a_1, \cdots, a_k) = \sum_{1 \leqq i < j \leqq k} \frac{1}{a_i a_j}.$$

Maybe this can be manipulated further.

17. Distributional properties of $\ell(X_N)$ for small N

Table XV exhibits the principal statistics of the distribution of the random variable $\ell_N = \ell(X_N)$ for $N \leqq 9$, calculated from Table XII.

TABLE XV

STATISTICS OF ℓ_N FOR $N \leqq 9$

N	$E(\ell_N)$	$E(\ell_N)/\sqrt{N}$	$\text{Var}\,\ell_N$	$\sqrt{(\text{Var}\,\ell_N)}$
1	1.00000	1.00000	0.00000	0.00000
2	1.50000	1.06066	0.25000	0.50000
3	2.00000	1.15470	0.33333	0.57735
4	2.41667	1.20830	0.41005	0.64035
5	2.79167	1.24844	0.49863	0.70614
6	3.14028	1.28201	0.57065	0.75541
7	3.46528	1.30975	0.63218	0.79510
8	3.77034	1.33302	0.69106	0.83130
9	4.05833	1.35278	0.74859	0.86521

I have tried extrapolating $E(\ell_N)/\sqrt{N}$ as $N \to \infty$ by calculating divided differences with $N^{-1/2}$ as argument; but this does not work at all well with the values of N available in Table XV, and yields to an estimate of c substantially less than 2.

18. Application to nonparametric testing of stationary sequences

As explained earlier, I took up Ulam's problem because it was a challenging mathematical problem, which aroused my curiosity. But, before long, of course, I asked myself if the mathematics might, by some happy accident, have applications.

One of the standard methods of testing whether a sequence of independent random variables is stationary, against the alternative that it has a trend, is to count the number of local maxima in the sequence. This test however has the drawback of being rather easily affected by local aberrations in the sequence: a test that took a more synoptic view of the whole sequence would be preferable. This situation is rather like that met in looking for periodicities in a stochastic process: the old-fashioned periodogram analysis suffers because genuine periodicities can be obscured by a few accidental phase-shifts; and the autocorrelation coefficient and its Fourier transform provide a better approach since they filter out these local irregularities.

The length of a longest ascending (or descending) subsequence should give quite a good synoptic nonparametric test statistic of stationarity. A sequence of length N will have a longest ascending subsequence of length about $2\sqrt{N}$ if it is stationary, but one of length proportional to N if it has an increasing trend. If the original sequence is reasonably long, so that N is much larger than $2\sqrt{N}$, the test will be very sensitive, especially because the Monte Carlo experiments suggest that ℓ_N has a small sampling variance (which might even be bounded as $N \to \infty$). The test statistic ℓ_N is also very easily computed by the algorithm in Section 9. But before the test can be put forward for practical use, we need to know more about Var ℓ_N.

19. Distribution of the number of ladder points

When I delivered the lecture on this paper at the Sixth Berkeley Symposium, Section 18 represented my thoughts on applications. But two days later Professor A. Dvoretzky, who had been in the audience, suggested to me that I should look into the corresponding nonparametric test based upon ladder points; he thought it likely that this would be both an easier mathematical problem and a more powerful test.

A point x_j in the sequence $X_N = \{x_1, x_2, \cdots, x_N\}$ is called a ladder point if $x_i \leqq x_j$ for all $i \leqq j$. What is the distribution of k_N, the number of ladder points in X_N, given that the elements of X are independently and identically distributed with a probability density function? Let

$$f_N(t) = \sum_{n=1}^{N} P(k_N = n)t^n \tag{19.1}$$

be the generating function of the distribution. Let us pass from X_N to X_{N+1}; then there is probability $1/(N+1)$ that x_{N+1} will add an extra ladder point to k_N, and probability $N/(N+1)$ that it will not. Hence

$$f_{N+1}(t) = f_N(t)(N+t)/(N+1); \tag{19.2}$$

whence

$$\begin{aligned} f_N(t) &= (1/N!)\prod_{r=0}^{N-1}(r+t) \\ &= (1/N!)\sum_{n=1}^{N}(-1)^{N+n}S_N^n t^n, \end{aligned} \tag{19.3}$$

where S_N^n are the Stirling numbers of the first kind ([17], p. 22). Hence

$$P(k_N = n) = (-1)^{N+n}S_N^n/N! = |S_N^n|/N! \tag{19.4}$$

gives the distribution of the number of ladder points. From (19.3) we can easily derive the mean and variance of k_N:

$$E(k_N) = \sum_{r=1}^{N} r^{-1} \sim \log N + \gamma, \tag{19.5}$$

$$\operatorname{Var}(k_N) = \sum_{r=1}^{N}(r-1)/r^2 \sim \log N + \gamma - \frac{\pi^2}{6}, \tag{19.6}$$

where γ is Euler's constant. For large N, we find that k_N has asymptotically a Poisson distribution with parameter $\log N$. Since $N^{-1}\log N$ is much smaller than $(2\sqrt{N})/N$, the test based on ladder points will be more sensitive than the one proposed in Section 18. It is one of those hard but sad facts of mathematics that the easier and less diverting mathematical problems are likely to be the more useful in practice. Now anything *both* useful *and* mathematically trivial will have been published several times over already; and one really ought to check the literature for references. I found papers by Chandler [2], Foster and Stuart [5], and Stuart [22], and—ironically enough—a couple of my own early papers [8], [9], which I had forgotten. (The conjectures in [9] were subsequently solved by Erdös [3] and by Moses and Wyman [20].)

20. Cross connections and conjectures

In Section 3 I said that it is better not to be influenced by reading the literature; and this should include forgetting about one's own work as well as ignoring other people's. Had I remembered my earlier work on Stirling's numbers, I would have been deprived of an important motive for thinking about Ulam's problem. Equally, it is difficult to escape from ideas and techniques that one has used before. The methods used in proving (8.15), in particular the introduction of a Poisson process and the associated Tauberian argument, originated from a paper [1] on the travelling salesman problem. There they sufficed for convergence with probability 1; but here, for various reasons, they only lead to convergence in probability. I want, of course, to show that (8.15) is also true with probability 1; and for this I believe that fresh ideas are needed, and I have tried hard to escape from the shackles of the earlier methods in [1], but without success. Preconceptions die hard.

But although the literature can be stultifying if one pays too much attention to proof, it can be stimulating if one concentrates upon conjectures. So I shall end by tracing some cross-connections between this paper and sundry problems and conjectures.

If N towns are distributed at random in a region of area A, and L_N is the length of the shortest journey that the travelling salesman must make to visit them all, then

$$L_N \sim C\sqrt{NA} \quad \text{as} \quad N \to \infty \tag{20.1}$$

with probability 1, where C is an absolute constant. The relations (8.15) and (20.1) are closely alike. In [1] it is shown that C satisfies various inequalities; but nobody has yet solved the problem of determining C exactly. What is wanted is theory for C along the lines of Sections 10, 11, and 12. There is also a similar problem for Steiner's network problem [11]. (Incidentally, I take this opportunity of correcting an error of calculation: in [1], p. 302, relation (7) should have read $2^{1/3}/3^{1/2} \leqq \alpha_3$, and consequently the relevant part of (8) should be $0.72742 \leqq \alpha_3$.)

I have written elsewhere [13] of the distinctions between "soft" and "hard" mathematics. One, though naturally not the only, distinction is that hard mathematics is often concerned with calculating the numerical value of a constant. To this extent Section 8 is soft mathematics dealing with generalities, while Sections 9 to 12 are hard mathematics aimed at determining the value of c. Of course, mathematical physicists are concerned with numerical values; and this is one of the reasons why their mathematical expertise tends to be sharper and stronger than that of pure mathematicians.

Rota's paper [21] has already featured in this story; and it has other cross connections which I shall mention briefly. He writes about the place of combinatorial analysis in mathematical research and he lists seven challenging problems: (i) the Ising problem; (ii) percolation theory; (iii) the number of necklaces, and Pólya's problem; (iv) self-avoiding random walks; (v) the travelling salesman problem; (vi) the coloring problem; and (vii) the pigeonhole principle and Ramsay's theorem. It is interesting to note how many of these topics are centered upon the determination of constants; and I am also pleased to find a high proportion of my own favourite problems in his list.

The Ising problem is one of the most celebrated problems in theoretical physics, nearly fifty years old now and still guarding its secrets about the numerical values of certain constants as well as more qualitative questions about the existence of singularities. Percolation theory is my own invention; and it has, as Rota explains, a close connection with the Ising problem. A general exposition together with a bibliography of percolation theory and its relation to the Ising problem and to various other questions in physics and chemistry appears in [6]: this also gives references to the self-avoiding walk problem, which is intimately connected with percolation theory. Another celebrated problem, closely associated with the Ising problem, is the monomer-dimer problem (see [14] for details and bibliography). This too asks for the value of a constant; but it also contains some "soft" mathematical problems, which would contribute greatly to our understanding of stochastic processes in more than one dimension if only we could solve them. The following problem is typical; and I am indebted to Professor David Blackwell for kindly supplying me with a translation of it into the language of modern mathematics.

"Denote by F the set of all functions f from the lattice points of the plane to $\{e, n, w, s\}$ such that

$$f(x, y) = w \Leftrightarrow f(x + 1, y) = e$$

and

$$f(x, y) = s \Leftrightarrow f(x, y + 1) = n.$$

For any finite set A of lattice points, denote by F_A the set of all restrictions of functions $f \in F$ to A. For any $B \supset A$, the uniform distribution on F_B induces a probability distribution $p(A, B)$ on F_A. Does $p(A, B)$ converge as B increases to the set of all lattice points, for every A?"

Percolation theory gave birth to subadditive stochastic processes, already discussed in Section 6. Despite considerable work on percolation problems, a great deal remains to be done. Rota [21] says that the percolation problem "was brilliantly solved by Michael Fisher, a British physicist now at Cornell University." This, however, is not quite correct, although Professor Fisher has done a great deal to advance our knowledge of these matters. I too was at one time under the misapprehension that the percolation problem was solved; in [6], p. 897, I wrote:

"Sykes and Essam have very recently (verbal communication to one of the authors) utilized somewhat similar conversions in a proof that the bond process critical probabilities of the triangular, square, and hexagonal lattices are respectively $2 \sin \pi/18$, $1/2$, and $1 - 2 \sin \pi/18$. Their brilliant solution of these three exceptionally difficult problems, all hitherto unsolved, is a most remarkable achievement."

Alas, these three problems are still unsolved: when Sykes and Essam published their work [23, 24], they wrote [24], p. 1125:

"We shall suppose, without offering a proof, that for real p $(0 \leqq p \leqq 1)$ the function K is singular at $p = p_c$, but nowhere else. This is to be expected in the light of exact results for closely related problems, and in particular, for percolation problems on lattices of the Bethe type for which K has been given exactly."

I am in no doubt that Sykes and Essam have got the right numerical answers (which agree, for example, with results obtained from series expansions and by Monte Carlo methods); and their exploitation of "matching" graphs is a valuable new tool in the subject; but they do not claim to have found a rigorous proof, and their argument rests upon plausible assumptions such as the one quoted above.

The travelling salesman problem we have already noted in (20.1). The most famous case of the coloring problem is to evaluate a constant, known to be either 4 or 5. Rota's illustration of the pigeonhole principle in [21] was discussed in Section 3. On self-avoiding walks, where one of the issues is to determine the numerical value of the so-called connective constant, Rota writes: "it is likely that this problem will be at least partly solved in the next few years, if interest in it stays alive." And this is the rub: he qualifies his decent optimism; for today an increasing number of graduate students, reared on the deficient diet of modern mathematics, take fright at difficult specific problems which they have neither the courage nor the intellectual training to tackle, and they turn aside to a tedious retilling of easier soils.

REFERENCES

[1] J. E. BEARDWOOD, J. H. HALTON, and J. M. HAMMERSLEY, "The shortest path through many points," *Proc. Cambridge Philos. Soc.*, Vol. 55 (1959), pp. 299–327.

[2] K. N. CHANDLER, "The distribution and frequency of record values," *J. Roy. Statist. Soc. Ser. B.*, Vol. 14 (1952), pp. 220–228.

[3] P. ERDÖS, "On a conjecture of Hammersley," *J. London Math. Soc.*, Vol. 28 (1953), pp. 232–236.

[4] W. FELLER, *An Introduction to Probability Theory and its Applications*, Vol. 2, New York, John Wiley, 1966, p. 225.

[5] F. G. FOSTER and A. STUART, "Distribution-free tests in time series based on the breaking of records," *J. Roy. Statist. Soc. Ser. B*, Vol. 16 (1954), pp. 1–22.

[6] H. L. FRISCH and J. M. HAMMERSLEY, "Percolation processes and related topics," *J. Soc. Indust. Appl. Math.*, Vol. 11 (1963), pp. 894–914.

[7] M. GARDNER, "Mathematical games," *Scientific American*, Vol. 216, (March 1967), pp. 123–129 and Vol. 216, (April 1967), pp. 116–123.

[8] J. M. HAMMERSLEY, "On estimating restricted parameters," *J. Roy. Statist. Soc. Ser. B*, Vol. 12 (1950), pp. 192–229.

[9] ———, "The sums of products of the natural numbers," *Proc. London Math. Soc. Ser. 3*, Vol. 1 (1951), pp. 426–452.

[10] ———, "A non-harmonic Fourier series," *Acta Math.*, Vol. 89 (1953), pp. 243–260.

[11] ———, "On Steiner's network problem," *Mathematika*, Vol. 8 (1961), pp. 131–132.

[12] ———, "A Monte Carlo solution of percolation in the cubic crystal," *Methods in Computational Physics*, New York, Academic Press, 1963, Vol. 1 pp. 281–298.

[13] ———, "The enfeeblement of mathematical skills by modern mathematics and by similar soft intellectual trash in schools and universities," *Bull. Inst. Math. Appl.*, Vol. 4 (1968), pp. 66–85.

[14] J. M. HAMMERSLEY and V. V. MENON, "A lower bound for the monomer-dimer problem," *J. Inst. Math. Appl.*, Vol. 6 (1970), pp. 341–364.

[15] J. M. HAMMERSLEY and D. J. A. WELSH, "First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory," *Bernoulli-Bayes-Laplace Anniversary Volume*, Berlin, Springer, 1965, pp. 61–110.

[16] G. H. HARDY, *Divergent Series*, Oxford, Oxford University Press, 1949.

[17] C. JORDAN, *Calculus of Finite Differences*, third edition. New York, Chelsea Publishing Company, 1965.

[18] J. F. C. KINGMAN, "The ergodic theory of subadditive stochastic processes." *J. Roy. Statist. Soc. Ser B*, Vol. 30 (1968), pp. 499–510.

[19] P. A. P. MORAN, "The random division of an interval," *J. Roy. Statist. Soc. Ser. B*, Vol. 9 (1947), pp. 92–98 and Vol 13 (1951), pp. 147–150.

[20] L. MOSES and M. WYMAN, "Asymptotic development of the Stirling numbers of the first kind," *J. London Math. Soc.*, Vol. 33 (1958), pp. 133–146.

[21] G.-C. ROTA, "Combinatorial analysis," *The Mathematical Sciences: Essays for the Committee on Support of Research in the Mathematical Sciences*, Cambridge, M.I.T. Press, 1969, pp. 197–208.

[22] A. STUART, "The efficiency of the records test for trend in normal regression," *J. Roy. Statist. Soc. Ser. B*, Vol. 19 (1957), pp. 149–153.

[23] M. F. SYKES and J. W. ESSAM, "Some exact critical percolation probabilities for bond and site problems in two dimensions," *Phys. Rev. Lett.*, Vol. 10 (1963), pp. 3–4.

[24] ———, "Exact critical percolation probabilities for site and bond problems in two dimensions," *J. Mathematical Phys.*, Vol. 5 (1964), pp. 1117–1127.

[25] S. M. ULAM, "Monte Carlo calculations in problems of mathematical physics," *Modern Mathematics for the Engineer: Second Series* (edited by E. F. Beckenbach) New York, McGraw-Hill, 1961.