INTRODUCTION

One aim of the theory of probability is the effective computation, perhaps only approximate, of probabilities that are given in principle. Of course there are other aims, for example the creation of an effective tool for thinking in a probabilistic way about physical problems and other applications, and also the development of an aesthetically satisfying theory that ties together the results of probabilistic computations for a particular class of structures, for example sums of independent random variables. Here I shall be concerned almost exclusively with a single approach to the approximate computation of probabilities and, more generally, expectations. This work may be thought of as an attempt to say something not entirely trivial about the approximate computation of expectations at an abstract level. I have tried, without complete success, to keep in mind the interaction of abstract ideas and concrete problems.

The problem of computing expectations, perhaps only approximately, can be divided into two parts. Let $(\Omega, \mathcal{B}, P)$ be a probability space and $E\colon \mathcal{X} \to R$ the expectation mapping associated with P on the linear space $\mathcal{X}$ of all real random variables $Z\colon \Omega \to R$ having finite expectations. In order to approximate EZ for such a Z we may first determine

(1) $$\ker E = \{Y\colon EY = 0\}$$

and then search $\ker E$ for a random variable that is approximately $Z-c$ for some constant c. We can then conclude that

(2) $$EZ \approx c.$$

In these notes I shall study these two subproblems from an abstract point of view and apply these considerations to a number of special problems. The basic abstract structure is described in the first seven pages of the first lecture, except for the proof of Lemma I.2 which is postponed until the end of the second lecture.

The principal systematic tool used here for the determination of ker E may be described as a method of auxiliary randomization. It may be instructive to indicate first how this can sometimes be used for the direct computation of probabilities. The probabilistic notation used will be somewhat informal. Let X denote a random point of a countable set Ω, distributed according to the probability measure P and let $W = \phi(X)$ be a discrete random variable. We can sometimes obtain the distribution of W in the following way. On a larger probability space we construct an exchangeable pair (X,X') of random points of Ω in such a way that the new X is still distributed according to P, and let $W' = \phi(X')$. Then (W,W') is an exchangeable pair, so that for any w and w'

$$\begin{aligned}(3)\qquad & P\{W = w\}P\{W' = w'|W = w\} \\ &= P\{W = w \;\&\; W' = w'\} = P\{W' = w \;\&\; W = w'\} \\ &= P\{W = w'\}P\{W' = w|W = w'\}.\end{aligned}$$

If we can compute or approximate the conditional probabilities in this equation we can solve for the ratio $P\{W = w'\}/P\{W = w\}$. If the system is connected in an obvious sense we can obtain the probabilities $P\{W = w\}$ (or approximations to them) from these ratios and the fact that the sum of the probabilities is 1. A simple example is provided by the binomial random variable $W = \sum_{i=1}^{n} X_i$ where the X_i are independent random variables satisfying $P\{X_i = 1\} = p$ and $P\{X_i = 0\} = 1-p$ with $0 < p < 1$. We choose I uniformly distributed in $\{1,\ldots,n\}$ independent of $X_1,\ldots,X_n$ and define $X_i' = X_i$ for $i \neq I$ but $X_I' = X^*$ (another random variable with the same distribution as $X_1,\ldots,X_n$ independent of $X_1\ldots X_n,I$). Then, with $W' = \sum_{i=1}^{n} X_i' = W + (X^*-X_I)$ the pair (W,W') of random variables is exchangeable and, for $w \in \{0,1,\ldots,n\}$,

(4) $$P\{W' = w - 1 | W = w\} = \frac{w}{n}(1-p),$$

and

(5) $$P\{W' = w + 1 | W = w\} = (1 - \frac{w}{n})p.$$

From these and (1) with $w' = w+1$ we obtain

(6) $$\frac{P\{W = w + 1\}}{P\{W = w\}} = \frac{(n - w)p}{(w + 1)(1 - p)}.$$

The binomial probabilities can easily be constructed from (6). Two slightly less trivial problems of this sort are discussed at the beginning of Lecture XII.

Such an exchangeable pair (X,X') can be used for the determination of ker E. Let $\mathfrak{F}$ be the space of antisymmetric functions $F: \Omega^2 \to R$ having finite expectation under the appropriate extension of P. Then, for such F,

(7) $$EF(X,X') = 0$$

and consequently, if we define $T: \mathfrak{F} \to \mathcal{X}$ by taking conditional expectation:

(8) $$(TF)(X) = E^X F(X,X')$$

we have

(9) $$E \circ T = 0.$$

In the case where Ω is finite, Lemma I.2 shows that, under an appropriate connectedness condition, which is clearly necessary,

(10) $$\ker E = \operatorname{im} T.$$

An analogous result must be true in infinite cases. In places I have used different ways to get at ker E. In particular, in the sixth lecture I have used integration by parts and, in the tenth lecture still another approach. These methods are related to the principal approach.

The second aspect of the problem of approximate computation of expectations, the search for a random variable in ker E approximately equal to Z-c with c constant when we want to approximate EZ, is handled by the development of an explicit formula (I.33) for the difference between the expectation of a random variable and its expectation under a specified approximating distribution, using the simple diagram, (I.28). This leads to the study of a certain linear mapping $U_0: \mathcal{X}_0 \to \mathfrak{F}_0$ associated with each approximating distribution, as described in (I.27)-(I.33). The easy analytic problems associated with U_0 are discussed for the univariate normal case in the first part of the second lecture, and for the Poisson case in the eighth lecture. In the latter case it is clear that special properties of the Poisson distribution have not been used very strongly. For the continuous case, the operator U_0 is studied, somewhat sketchily, in a fairly general context, in the sixth lecture. The basic idea is applied to the normal approximation problem in several of the lectures, as indicated in the table of contents. In the discrete case, only binomial and Poisson approximation are considered.

The method described in this series of lectures for deriving a simple explicit expression for the difference between the true expectation and its approximation by a simpler expectation, for example normal or Poisson, seems to be as successful as one could wish but the problem of bounding or approximating that difference has not really been solved. At least three different attitudes toward this problem are possible:

(i) This remainder is an expectation, so it is reasonable to try to apply the same method that was successful for the original problem.

(ii) One can try to bound the remainder by special devices in each case.

(iii) The notion of approximate independence of two random variables seems to arise frequently in bounding or approximating the remainder. To approximate E XY we may define a random variable Z, together with X and Y, on a larger probability space, in such a way that Z is independent of X and also Z is nearly equal to Y and then use

$$E\,XY = E\,XZ + E\,X(Y - Z) \qquad (11)$$

$$= (E\,X)(E\,Z) + E\,X(Y - Z).$$

The first approach is successful in the treatment in Lecture VIII of the Poisson approximation to the distribution of the number of successes in a large number of independent trials. There the remainder in the basic identity (VIII.27) is bounded rather trivially in (VIII.43). The remainder can then be evaluated by another application of the basic identity, leading to the inequality (VIII.47). It is likely that this approach will be successful in other cases but there are difficulties. The remainder is not eliminated but rather replaced by another remainder, which we hope is substantially smaller. Thus some additional device must be applied. I have been completely unable to apply this approach in the abstract context of (I.31) (or XV.5). However I still believe that, properly interpreted, this approach will be useful.

The third approach is best illustrated by Lecture XI, which is the second lecture on counting Latin rectangles. In order to study some properties of a random permutation Π of $\{1,...,n\}$ we first use auxiliary randomization in the form of a random ordered pair (I,J) of distinct elements of $\{1,...,n\}$ independent of Π. This yields the basic identity (VII.36) of the earlier lecture on this subject. From this a crude bound is obtained in Lecture VII. This is substantially improved in Lecture XI by observing that the two factors under the expectation sign in the remainder of (VII.36) are nearly independent. The construction used to prove this suggests some similarity between this and the first approach. An exchangeable pair $((\Pi,I,J), (\Pi'',I'',J''))$ is constructed in which Π'' differs only slightly from Π. This construction should also be useful in other problems concerning a random permutation such as the study of the distribution of the sum of a random diagonal discussed in Lecture III.

The rather weak dependence among the first fourteen lectures is indicated in Fig. 1. The fifteenth lecture makes some reference to material of nearly all the other lectures.

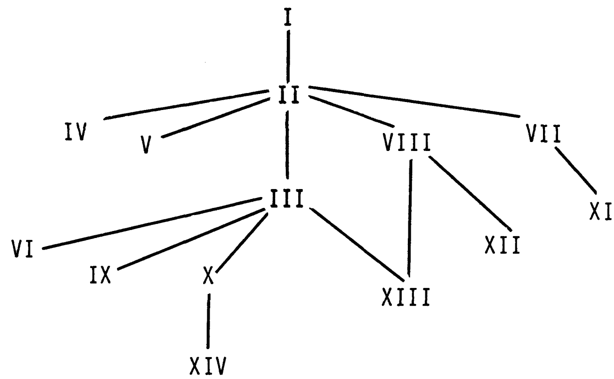

Fig. 1

Practically none of the concrete results are new. The basic approach seems to have been new at the time of my paper at the Sixth Berkeley Symposium. The algebraic formalism of (I.33), (XIV.33), and (XV.5) is new. Lectures VII and XI on counting Latin rectangles are a corrected version (I hope) of my 1978 paper in the Journal of Combinatorial Theory (Series A). Lecture VIII on Poisson approximation is based on the work of Louis Chen and Lecture XIII on random graphs is based on work of Barbour and Eagleson. Lecture IV is based on joint work of Diaconis and the author, reported in Diaconis (1977). Some applications of the method by other people are listed in the bibliography and, in some cases, described briefly in Lecture XV.

I believe that this approach will eventually turn out to be quite powerful but I recognize that I have not yet made a convincing case for this statement. I regret that the treatment of many special problems and even of the basic idea is poorly organized and incomplete. Some special problems are treated more thoroughly by other authors in papers listed in the bibliography. I hope that my emphasis, systematic in intent though somewhat disorganized, on the basic ideas will help people think about this class of problems more effectively. Further delay in getting out these notes, in the hope of improving them, seems unwise.

I am grateful to the National Science Foundation for its support of this work at the Mathematical Sciences Research Institute in Berkeley, at

Stanford University, and elsewhere. I thank Michael Perlman and the University of Washington at Seattle for their invitation to spend the winter and spring of 1982 in Seattle, where some of this work was done. Persi Diaconis suggested that I work on some of the special problems treated here, and Akimichi Takemura, Louis Chen, and others have made very helpful comments. I am grateful to my wife Margaret for her encouragement and for her typing of the penultimate version of this manuscript. I also thank Norma Lucas for typing the final version.