LECTURE XV. SUMMARY

In these concluding remarks I shall summarize the basic ideas of this series of lectures and raise a few questions about possible future work. In particular I shall try to emphasize the desirability of carrying out this work at various levels of abstraction, especially the lowest and the highest. I shall also make a few comments on the bibliography.

Let us recall the basic diagram, as it was presented in the fourteenth lecture:

(1)
$$\begin{tikzcd}
\mathfrak{F} \arrow[r, "T"] & \mathcal{X} \arrow[r, "E"] & \mathcal{Y} \\
\mathfrak{F}_0 \arrow[u, "\alpha"] \arrow[r, shift left, "T_0"] & \mathcal{X}_0 \arrow[u, "\beta"] \arrow[l, shift left, "U_0"] \arrow[r, shift left, "E_0"] & \mathcal{Y}_0 \arrow[u, "\gamma"'] \arrow[l, shift left, "\iota_0"]
\end{tikzcd}$$

The linear mappings in this diagram are assumed to satisfy the following conditions:

(2) $$E \circ T = 0$$

(3) $$\gamma = E \circ \beta \circ \iota_0$$

(4) $$I_{\mathcal{X}_0} = \iota_0 \circ E_0 + T_0 \circ U_0 .$$

From these assumptions we easily concluded that

(5) $$E \circ \beta - \gamma \circ E_0 = E \circ (\beta \circ T_0 - T \circ \alpha) \circ U_0 .$$

Formally this expresses the departure from commutativity of the right-hand square in terms of the corresponding expression for the left-hand square. Typically we introduce this diagram in order to approximate the restriction of E: $\mathcal{X} \to \mathcal{Y}$ to a subspace $\mathcal{X}_0$, with β: $\mathcal{X}_0 \to \mathcal{X}$ the appropriate inclusion mapping. Then (5) may express $E \circ \beta$ in terms of a part, $\gamma \circ E_0$, that is easy to compute, and the right-hand side, which is small if the construction has been successful.

A number of questions can be raised at this level of abstraction. How broadly is this approach applicable? Perhaps it has already been applied in other fields, independently of this work. Can the identity (5) be applied iteratively? In the abstract formulation, condition (3) seems to make this difficult. However, in special cases, iterative application is possible, as in (VIII.30), which is the result of two applications of (VIII.27). Can this be understood within the abstract formulation? It is important to study in detail the lower row of diagram (1) in many special cases. This was done in Lecture II for a special case associated with univariate normal approximation problems and, in Lecture VIII for a special case associated with Poisson approximation problems. The binomial case has been treated very superfically. In Lecture VI there is also a suggestions of a moderately general treatment of the case where T_0 is a first-order ordinary differential operator. It would be desirable to give a more thorough treatment of all of these cases. Two other cases that seem likely to introduce interesting new features but no insuperable difficulties are infinitely divisible laws and the multivariate normal case.

Now let us return to diagram (1), in particular its top row, in the case that interests us most, where E is an expectation operator. More precisely, $\mathcal{X}$ is the linear space of real random variables on a probability space $(\Omega, \mathfrak{B}, P)$ having finite expectation and E: $\mathcal{X} \to R(=\mathcal{Y})$ is expectation. In most of this work the mapping T: $\mathfrak{F} \to \mathcal{X}$ has been defined in the following way. On an enlarged probability space we construct an exchangeable pair of random objects (X,X') where X, taking values in Ω, is distributed according to P, and $\mathfrak{F}$ is the space of antisymmetric functions of (X,X') having finite expectation. Then T: $\mathfrak{F} \to \mathcal{X}$

is defined to be E^X, the operation of conditional expectation with TF for $F \in \mathfrak{F}$ interpreted in the obvious way as an element of $\mathfrak{X}$. For more details see the first lecture. In Lemma I.2, whose proof is postponed until the second lecture, it is pointed out that, in the case of finite Ω, (2) is strengthened to

$$\ker E = \operatorname{im} T, \tag{6}$$

provided the exchangeable pair (X,X') connects Ω in an appropriate sense. Presumably an analogous result holds even if Ω is infinite.

It seems natural to ask whether there are other useful ways of constructing $\mathfrak{F}$ and $T\colon \mathfrak{F} \to \mathfrak{X}$ to complete the top row of the basic diagram. In the very special case of a sum of independent random variables, this is done in the sixth lecture using integration by parts. Unfortunately I have not had time, in this series of lectures, to go into the fact that this use of integration by parts is a limiting case of the earlier approach.

In parts of these lectures I have not used the full formalism of the first lecture. Instead I have used the exchangeable pair (X,X') for a direct computation of the ratio of the probabilities of two different values of a statistic W:

$$\frac{P\{W=w_2\}}{P\{W=w_1\}} = \frac{P\{W'=w_2|W=w_1\}}{P\{W'=w_1|W=w_2\}}, \tag{7}$$

where W is a certain function of X and W' the corresponding function of X'. This was applied in Lecture XII to the computation of the exact distribution in two combinatorial problems. Of course the exchangeable pair (X,X') must be chosen in such a way that the right-hand side of (7) is easy to compute or to approximate. There is another idea that is intuitively appealing, although I have not had any real success with it. We can construct a reversible stationary Markov chain $X_1,X_2,\ldots$ with (X_i,X_{i+1}) having the same joint distribution as (X,X'). Then it may be possible to bring to bear some results concerning stationary distributions of Markov chains, at least in an intuitive way.

I shall conclude with a few remarks about the literature. Some aspects of this method, introduced in my paper in the Proceedings of the Sixth Berkeley Symposium, were developed earlier by other people. The reflection principle in the study of simple random walk or Brownian motion is an example of the introduction of an exchangeable pair to obtain a trivial solution of an otherwise non-trivial problem. The approach of Metropolis (1953) to Monte Carlo calculations may also be a precursor of this method. A recent paper on this subject is Gidas (1985). The jackknife estimate of the variance and the influence function in robust statistics also seem to be related to the present method.

In my paper in the Proceedings of the Sixth Berkeley Symposium, I tried to derive an abstract normal approximation theorem in which the bound for the error is, roughly speaking, of the right order of magnitude, under favorable conditions. I had hoped to rewrite that paper for the present set of notes, but did not succeed. Because of poor execution of some details it was necessary to assume the existence of eighth moments. In addition there are a number of minor errors and misprints. In particular it is clear from the comments below (3.33) that the bound given there for the error in the normal approximation when the dependence decreases exponentially should be $An^{-\frac{1}{2}} \log^2 n$, rather than $An^{-\frac{1}{2}}$. It is now known that this is not the best possible bound. I believe the results are essentially correct, but this paper and the related papers of Tikhomirov (1980) and Takahata (1981, 1983) are not easy to read.

Tikhomirov (1980) obtained somewhat better results in special cases by a modification of this method. Essentially, he obtained a differential equation for the characteristic function of a sum of random variables, under certain conditions, by applying a variant of the basic identity (X.3) to a complex exponential function. After solving this equation he could apply the usual Fourier inversion techniques. In this way he obtained better bounds in the case of exponentially decreasing dependence under weaker conditions (third and, for some results, fourth moments). He also obtained a reasonable explicit bound for the error in the m-dependent case.

Takahata (1981, 1983) obtained results on the approximate normality of sums of dependent random variables that are, in some ways, better than the results of Tikhomirov. In particular he considered random fields rather than only random sequences. Earlier Erickson (1974), by very simple arguments, had obtained some interesting results on L_1 bounds for the error in normal approximations.

Barbour and Hall (1984a, 1984c) made an ingenious application of this method to derive (essentially) a lower bound for the maximum error of the normal approximation to the distribution of a sum of independent random variables. Let $Y_1,\ldots,Y_n$ be independent random variables with mean 0 and $EY_i^2 = \sigma_i^2$ with $EW^2 = 1$, where $W = \sum Y_i$. Let

$$\delta = \sum E\{Y_i^2 \mathcal{I}[|Y_i|>1]\} + \sum E\{Y_i^4 \mathcal{I}[|Y_i|\leq 1]\} + |\sum E\{Y_i^3 \mathcal{I}[|Y_i|\leq 1]\}|$$

and let

$$\Delta = \sup_w |P\{W \leq w\} - \Phi(w)\}|.$$

Then there is an absolute constant C such that

$$\delta < C(\Delta + \sum \sigma_i^4).$$

In 1984a they also gave a simple proof of a slight improvement on the Berry-Esseen Theorem.

My former student, Louis Chen, developed the theory in many important ways. In Chen (1975a) he studied the Poisson approximation for the distribution of a sum of dependent random variables taking on only the values 0 and 1. My Lecture VIII is essentially the specialization of this argument to the independent case, with a small improvement due to Barbour and Eagleson. He and his students and coworkers have also obtained new results on normal approximation. In Chen (1979) he gave a general survey of the field.

It seems likely to be useful to apply these methods to the study of concrete problems possessing a simple structure, perhaps based on symmetry or independence, but not falling under any of the standard general theorems. Examples are the work of the author and Diaconis (1977), described in Lecture IV, and that of Barbour (1982) on random graphs, part of which was presented in Lecture XIII. Another example is the work of the author on counting Latin rectangles, given in Lectures VII and XI. It would be desirable to carry this work further, in a serious attempt to obtain detailed new results rather than merely to illustrate the method.

Now let us turn to the question of possible directions for future work. First there are the obvious possibilities of continuing essentially in the spirit of the present work. Some of the topics that come to mind are

a) an abstract Berry-Esseen theorem

b) multivariate normal approximation

c) infinitely divisible distributions, in particular the stable laws

d) the discrete analogue of Lecture VI

e) a serious study of large deviations

f) some further illustrative problems.

In the present notes, some fairly concrete problems have been considered in the fourth, seventh, eleventh, twelfth and thirteenth lectures as illustrations of the theory.

It would be interesting to try to apply the methods of these lectures to the study of special problems with the primary aim of obtaining useful results and exploring a whole field of applications rather than merely illustrating the general theory. Some possible fields are

a) random allocations, in the sense of the book of that title by Kolchin, Sevastyanov, and Chistyakov (1978)

b) random permutations (an introduction to this subject is contained in the above-mentioned book)

c) random graphs

d) robust regression

e) probabilistic number theory.

The first three of these have been discussed briefly in these notes as illustrations of the general theory but without any serious attempt at obtaining a broad view of any of these fields. In particular it would be desirable to explore iterated application of the basic identities, approximations that do not take the usual form, such as normal or Poisson, and the possibility of combining these ideas with Monte-Carlo methods.

Perhaps it would be useful to practice for the program of the last paragraph by obtaining careful detailed solutions in usable form for some very special, simple problems. Some examples are

a) the distribution of the Wilcoxon rank-sum statistic (see, for example, Lehmann (1975))

b) a sum of independent identically distributed random variables taking on only three possible values, starting with the case where these are in arithmetic progression

c) The number of successes in n independent trials, not necessarily with identical probabilities, including improvements on the normal approximation and large deviation results.

Of course we should not expect to obtain new results for these problems, but we could aim for simplicity and improved understanding of the basic tools.