PREFACE

Probability theory in the first half of the twentieth century was substantially dominated by the formulation and proof of the classical limit theorems — laws of large numbers, central limit theorem, law of the iterated logarithm — for sums of independent random variables. The central limit theorem in particular has found regular application in statistics, and forms the basis of the distribution theory of many test statistics. However, the classical approach to the CLT relied heavily on Fourier methods, which are not naturally suited to providing estimates of the accuracy of limits such as the CLT as approximations in pre-limiting circumstances, and it was only in 1940 that Berry and Esseen, by means of the smoothing inequality, first obtained the correct rate of approximation in the form of an explicit, universal bound. Curiously enough, the comparable theorem for the conceptually simpler Poisson law of small numbers was not proved until 26 years later, by Le Cam.

These theorems all concerned sums of independent random variables. However, dependence is the rule rather than the exception in applications, and had been increasingly studied since 1950. Without independence, Fourier methods are much more difficult to apply, and bounds for the accuracy of approximations become correspondingly more difficult to find; even for such frequently occurring settings as sums of stationary, mixing random variables or the combinatorial CLT, establishing good rates seemed to be intractable.

It was into this situation that Charles Stein introduced his startling technique for normal approximation. Now known simply as Stein's method, the technique relies on an indirect approach, involving a differential operator and a cleverly chosen exchangeable pair of random variables, which are combined in almost magical fashion to deliver explicit estimates of approximation error, with or without independence. This latter feature, in particular, has led to the wide range of application of the method.

Stein originally developed his method to provide a new proof of the combinatorial CLT for use in a lecture course, and its first published application, in the *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* in 1972, was to give bounds for the accuracy of the CLT for sums of stationary, mixing random variables. Since then, the scope of his discovery has expanded rapidly. Poisson approximation was studied in 1975; the correct Lyapounov error bound in the combinatorial CLT was obtained in 1984; the method was extended to the approximation of the distributions of whole random processes in 1988; its importance in the theoretical underpinning of molecular sequence comparison algorithms was recognized in 1989; rates of convergence in the multivariate CLT were derived in 1991; good general bounds in the multivariate CLT, when dependence is expressed in terms of neighborhoods of possibly very general structure, were given in 1996; and Stein's idea of arguing by way of a concentration inequality was developed in 2001 to a point where it can be put to very effective use.

Despite the progress made over the last 30 years, the reasons for the effectiveness of Stein's method still remain something of a mystery. There are still many open problems, even at a rather basic level. Controlling the behavior of the solutions of the Stein equation, fundamental to the success of the method, is at present a difficult task, if the probabilistic approach cannot be used. The field of multivariate discrete distributions is almost untouched. There is a relative of the concentration inequality approach, involving the comparison of a distribution with its translations, which promises much, but is at present in its early stages. Point process approximation, other than in the Poisson context, is largely unexplored: the list goes on.

Due to its broad range of application, Stein's method has become particularly important, not only in the future development of probability theory, but also in a wide range of other fields, some theoretical, some extremely practical. These include spatial statistics, computer science, the theory of random graphs, computational molecular biology, interacting particle systems, the bootstrap, the mathematical theory of epidemics, algebraic analogues of probabilistic number theory, insurance and financial mathematics, population ecology and the combinatorics of logarithmic structures. Many, in their turn, because of their particular structure, have led to the development of variants of Stein's original approach, with their own theoretical importance, one such being the coupling method.

This volume contains an introduction to Stein's method in four chapters, corresponding to the tutorial lectures given during the meeting

STEIN'S METHOD AND APPLICATIONS:
A PROGRAM IN HONOR OF CHARLES STEIN,

held in Singapore at the Institute for Mathematical Sciences, from 28 July to 31 August 2003. The material provides a detailed introduction to the theory and application of Stein's method, in a form suitable for graduate students who want to acquaint themselves with the method. The accompanying volume, consisting of papers given at the workshop held during the same meeting, provides a cross-section of the research work currently being undertaken in this area.

To get a flavour of the magic and mystery of Stein's method, take the following elementary setting: $X_1, X_2, \ldots, X_n$ are independent 0–1 random variables, with $\mathbb{P}[X_i = 1] = 1 - \mathbb{P}[X_i = 0] = p_i$, and W denote their sum. How close is the distribution $\mathcal{L}(W)$ to the Poisson distribution $\mathrm{Po}\,(\lambda)$ with mean $\lambda = \sum_{i=1}^n p_i$? A good answer can be obtained in three small steps.

(1) For any $A \subset \mathbb{Z}_+$, recursively define the function $g = g_{\lambda,A}$ on $\mathbb{Z}_+$ by setting $g(0) = 0$ and then

$$\lambda g(j+1) = jg(j) + \mathbf{1}_A(j) - \mathrm{Po}\,(\lambda)\{A\} \tag{0.1}$$

for $j = 0, 1, 2, \ldots$. Then, by taking expectations, it follows that

$$\mathbb{P}[W \in A] - \mathrm{Po}\,(\lambda)\{A\} = \mathbb{E}\{\lambda g(W+1) - Wg(W)\}, \tag{0.2}$$

as long as $jg(j)$ is bounded in j (it is).

(2) Then note that $X_i g(W) = X_i g(W_i + 1)$, where $W_i = \sum_{j \neq i} X_j$, because X_i takes only the values 0 and 1. Since also W_i is *independent* of X_i, it thus follows that $\mathbb{E}\{X_i g(W)\} = p_i \mathbb{E} g(W_i + 1)$, and hence that

$$\mathbb{E}\{Wg(W)\} = \sum_{i=1}^n p_i \mathbb{E} g(W_i + 1). \tag{0.3}$$

(3) Combining (0.2) and (0.3), we have

$$\begin{aligned} |\mathbb{P}[W \in A] - \mathrm{Po}(\lambda)\{A\}| &= \left| \sum_{i=1}^n p_i \mathbb{E}[g(W+1) - g(W_i+1)] \right| \\ &= \left| \sum_{i=1}^n p_i \mathbb{E}[g(W_i + X_i + 1) - g(W_i+1)] \right|, \end{aligned}$$

from which it follows that

$$|\mathbb{P}[W \in A] - \mathrm{Po}\,(\lambda)\{A\}| \le k(\lambda) \sum_{i=1}^{n} p_i^2 \tag{0.4}$$

for all $A \subset \mathbb{Z}_+$, where

$$k(\lambda) := \sup_{A \subset \mathbb{Z}_+} \sup_{j \ge 1} |g_{\lambda,A}(j+1) - g_{\lambda,A}(j)|,$$

since X_i differs from 0 only with probability p_i, when it takes the value 1. And it can also be shown that $k(\lambda) \le (1 - e^{-\lambda})/\lambda$.

The upshot of this argument is that the difference between the probability given by $\mathcal{L}(W)$ to *any* set A and that assigned to it by $\mathrm{Po}\,(\lambda)$ is at most

$$\lambda^{-1}(1 - e^{-\lambda}) \sum_{i=1}^{n} p_i^2 \le \max_{1 \le i \le n} p_i, \tag{0.5}$$

a remarkably neat and surprisingly sharp result. This volume shows how the simple argument that led to it (the Stein–Chen method) fits into the much more general and powerful framework of Stein's method. Reasons are advanced for choosing equation (0.1) in connection with the Poisson distribution $\mathrm{Po}\,(\lambda)$. Some rules are given for constructing analogous equations for other distributions, both on the line and on more elaborate spaces, such as measure spaces, and some help is also provided with bounding the counterparts of $k(\lambda)$ that emerge. Finally, ways of modifying (0.3) when W is a sum of dependent random elements are also proposed.

The material is arranged in four chapters, successively addressing the normal distribution, Poisson and compound Poisson distributions, Poisson point processes, and then quite general distributions. Each chapter is written by an expert in the field. We hope that the resulting tutorial survey will encourage the reader to become as enthusiastic about Stein's method as we are.

December 2004

A. D. Barbour
Louis H. Y. Chen
Program Co-Chairs