Chapter 1
Introduction

1.1 The Central Limit Theorem

The Central Limit Theorem is one of the most striking and useful results in probability and statistics, and explains why the normal distribution appears in areas as diverse as gambling, measurement error, sampling, and statistical mechanics. In essence, the Central Limit Theorem in its classical form states that a normal approximation applies to the distribution of quantities that can be modeled as the sum of many independent contributions, all of which are roughly the same size.

Thus mathematically justified, at least asymptotically, in practice the normal law may be used to approximate quantities ranging from a p-value of a hypothesis tests, the probability that a manufacturing process will remain in control or the chance of observing an unusual conductance reading in a laboratory experiment. However, even though in practice sample sizes may be large, or may appear to be sufficient for the purposes at hand, depending on that and other factors, the normal approximation may or may not be accurate. It is here the need for the evaluation of the quality of the normal approximation arises, which is the topic of this book.

The seeds of the Central Limit Theorem, or CLT, lie in the work of Abraham de Moivre, who, around the year 1733, not being able to secure himself an academic appointment, supported himself consulting on problems of probability and gambling. He approximated the limiting probabilities of the binomial distribution, the one which governs the behavior of the number

$$S_n = X_1 + \cdots + X_n \tag{1.1}$$

of successes in an experiment which consists of n independent Bernoulli trials, each one having the same probability $p \in (0, 1)$ of success. de Moivre realized that even though the sum

$$P(S_n \le m) = \sum_{k \le m} \binom{n}{k} p^k (1-p)^{n-k}$$

that yields the cumulative probability of m or fewer successes becomes unwieldy for even moderate values of n, there exists an easily computable, normal approximation to such probabilities that can be quite accurate even for moderate values of n.

L.H.Y. Chen et al., *Normal Approximation by Stein's Method*,
Probability and Its Applications,
DOI 10.1007/978-3-642-15007-4_1,

Only many years later with the work of Laplace around 1820 did it begin to be systematically realized that the normal limit holds in much greater generality. The result was the classical Central Limit Theorem, which states that $W_n \to_d Z$, that is, W_n converges in distribution to Z, whenever

$$W_n = (S_n - n\mu)/\sqrt{n\sigma^2} \tag{1.2}$$

is the standardization of a sum S_n, as in (1.1), of independent and identically distributed random variables each with mean μ and variance σ^2. Here, Z denotes a standard normal variable, that is, one with distribution function $P(Z \le x) = \Phi(x)$ given by

$$\Phi(x) = \int_{-\infty}^{x} \varphi(u)\,du \quad \text{where } \varphi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right),$$

and we say a sequence of random variables Y_n is said to converge in distribution to Y, written $Y_n \to_d Y$, if

$$\lim_{n\to\infty} P(Y_n \le x) = P(Y \le x) \quad \text{for all continuity points } x \text{ of } P(Y \le x). \tag{1.3}$$

Generalizing further, but still keeping the variables independent, the question of when a sum of independent but not necessarily identically distributed random variables is asymptotically normal is essentially completely answered by the Lindeberg–Feller–Lévy Theorem (see Feller 1968b), which shows that the Lindeberg condition is sufficient, and nearly necessary, for the normal limit to hold. For a more detailed, and delightful account of the history of the CLT, we refer the reader to LeCam (1986).

When the quantity W_n given by (1.2) is a normalized sum of i.i.d. variables $X_1, \ldots, X_n$ with finite third moment, the works of Berry (1941) and Esseen (1942) were the first to give a bound on the normal approximation error, in terms of some universal constant C, of the form

$$\sup_{z\in\mathbb{R}} \big| P(W_n \le z) - P(Z \le z) \big| \le \frac{C E|X_1|^3}{\sqrt{n}}.$$

This prototype bound has since been well studied, generalized and applied in practice, and it appears in many guises in the pages that follows. Esseen's original upper bound on C of magnitude 7.59 has been markedly decreased over the years, the record currently now held by Tyurin (2010) who proved $C \le 0.4785$.

With the independent case tending toward resolution, attention can now turn to situations where the variables exhibit dependence. However, as there are countless ways variables can fail to be independent, no single technique can be used to address all situations, and no theorem parallel to the Lindeberg–Feller–Lévy theorem is ever to be expected in this greater generality. Consequently, the literature for validating the normal approximation in the presence of dependence now fragments somewhat into various techniques which can handle certain specific structures, or assumptions, two notable examples being central limit theorems proved under mixing conditions, and those results that can be applied to martingales.

Characteristic function methods have proved essential in making progress in the analysis of dependence, and though they are quite powerful, they rely on handling distributions through their transforms. In doing so it is doubtless that some probabilistic intuition is lost. In essence, the Stein method replaces the complex valued characteristic function with a real characterizing equation through which the random variable, in its original domain, may be manipulated, and in particular, coupled.

1.2 A Brief History of Stein's Method

Stein's method for normal approximation made its first appearance in the ground breaking work of Stein (1972), and it was here that the characterization of the normal distribution on which this book is based was first presented. That is, the fact that $Z \sim \mathcal{N}(0, \sigma^2)$ if and only if

$$E\big[Zf(Z)\big] = \sigma^2 E\big[f'(Z)\big], \tag{1.4}$$

for all absolutely continuous functions f for which the above expectations exist. Very soon thereafter the work of Chen (1975) followed, applying the characterizing equation method to the Poisson distribution based on the parallel fact that $X \sim \mathcal{P}(\lambda)$, a Poisson variable with parameter λ, if and only if

$$E\big[Zf(Z)\big] = \lambda E\big[f(Z+1)\big],$$

for all functions f for which the expectations above exist. From this point it seemed to take a number of years for the power of the method in both the normal and Poisson cases to become fully recognized; for Poisson approximation using Stein's method, see, for instance, the work of Arratia et al. (1989), and Barbour et al. (1992). The key identity (1.4) for the normal was, however, put to good use in the meantime.

In another landmark paper, Stein (1981) applied the characterization that he had proved earlier for the purpose of normal approximation to derive minimax estimates for the mean of a multivariate normal distribution in dimensions three or larger. In particular, he shows, using the multivariate version of (1.4), that when $\mathbf{X}$ has the normal distribution with mean $\boldsymbol{\theta}$ and identity covariance matrix, then the mean squared error risk of the estimate $\mathbf{X} + g(\mathbf{X})$, for an almost everywhere differentiable function $g : \mathbb{R}^p \to \mathbb{R}^p$, is unbiasedly estimated by $p + \|g(\mathbf{X})\|^2 + 2\nabla \cdot g(\mathbf{X})$. This 1981 work builds on the earlier and rather remarkable and surprising result of Stein (1956), that shows that the usual sample mean estimate $\overline{\mathbf{X}}$ for the true mean $\boldsymbol{\theta}$ of a multivariate normal distribution $\mathcal{N}_p(\boldsymbol{\theta}, \mathbf{I})$ is not admissible in dimensions three and greater; the multivariate normal characterization given in Stein (1981) provides a rather streamlined proof of this very counterintuitive fact.

Returning to normal approximation, by 1986 Stein's method was sufficiently cohesive that its foundations and some illustrative examples could be laid out in the manuscript of Stein (1986), with the exchangeable pair approach being one notable cornerstone. This manuscript also considers approximations using the binomial and the Poisson, and other probability estimates related to but not directly concerning the normal. In the realm of normal approximation, this work rather convincingly

demonstrated the potential of the method under dependence by showing how it could be used to assess the quality of approximations for the distribution of the number of empty cells in an allocation model, and the number of isolated trees in the Erdös–Rényi random graph. For a personal history up to this time from the view point of Charles Stein, see his recollections in DeGroot (1986).

The period following the publication of Stein's 1986 manuscript saw a veritable explosion in the number of ideas and applications in the area, a fact well illustrated by the wide range of topics covered here, as well as in the two volumes of Barbour and Chen (2005b, 2005c), and those referred to in the bibliographies thereof. Including up to the present day, powerful extensions and applications of the method continue to be discovered that were, at the time of its invention, completely unanticipated.

1.3 The Basic Idea of Stein's Method

To show a random variable W has a distribution close to that of a target distribution, say that of the random variable Z, one can compare the values of the expectations of the two distributions on some class of functions. For instance, one can compare the characteristic function $\phi(u) = Ee^{iuW}$ of W to that of Z, thus encapsulating all expectations of the family of functions e^{iuz} for $u \in \mathbb{R}$. And indeed, as this family of functions is rich enough, closeness of the characteristic functions implies closeness of the distributions. When studying the sum of random variables, and independent random variables in particular, the characteristic function is a natural choice, as convolution in the space of measures become products in the realm of characteristic functions. Powerful as they may be, one may lose contact with probabilistic intuition when handling complex functions in the transform domain. Stein's method, based instead on a direct, random variable characterization of a distribution, allows the manipulation of the distribution through constructions involving the basic random quantities of which W is composed, and coupling can begin to play a large role.

Consider, then, testing for the closeness of the distributions of W and Z by evaluating the difference between the expectations $Eh(W)$ and $Eh(Z)$ over some collection of functions h. At first there appears to be no handle that we can apply, the task as stated being perhaps overly general. Nevertheless, it seems clear that if the distribution of W is close to the distribution of Z then the difference $Eh(W) - Eh(Z)$ should be small for many functions h. Specializing the problem, for a specific distribution, we may evaluate the difference by relying on a characterization of Z. For instance, by (1.4), the distribution of a random variable Z is $\mathcal{N}(0, 1)$ if and only if

$$E\big(f'(Z) - Zf(Z)\big) = 0 \tag{1.5}$$

for all absolutely continuous functions f for which the expectation above exists. Again, if the distribution of W is close to that of Z, then evaluating the left hand side of (1.5) when Z is replaced by W should result in something small. Putting these two differences together, from the Stein characterization (1.5) we arrive at the Stein equation

$$f'(w) - wf(w) = h(w) - Eh(Z). \tag{1.6}$$

Now, given h, one solves (1.6) for f, evaluates the left hand side of (1.6) at W and takes the expectation, obtaining $Eh(W) - Eh(Z)$.

Perhaps at first glance the problem has not been made any easier, as the evaluation of $Eh(W) - Eh(Z)$ has been replaced by the need to compute $E(f'(W) - Wf(W))$. Yet the form of what is required to evaluate is based on the normal characterization, and, somehow, for this reason, the expectation lends itself to calculation for W for which approximation by the normal is appropriate. Borrowing, essentially, the following 'leave one out' idea from Stein's original 1972 paper, let $\xi_1, \ldots, \xi_n$ be independent mean zero random variables with variances $\sigma_1^2, \ldots, \sigma_n^2$ summing to one, and set

$$W = \sum_{i=1}^{n} \xi_i.$$

Then, with $W^{(i)} = W - \xi_i$, for some given f, we have

$$E\big(Wf(W)\big) = E\sum_{i=1}^{n} \xi_i f(W) = E\sum_{i=1}^{n} \xi_i f\big(W^{(i)} + \xi_i\big).$$

If f is differentiable, then the summand may be expanded as

$$\xi_i f\big(W^{(i)} + \xi_i\big) = \xi_i f\big(W^{(i)}\big) + \xi_i^2 \int_0^1 f'\big(W^{(i)} + u\xi_i\big)\, du,$$

and, since $W^{(i)}$ and ξ_i are independent, the first term on the right hand side vanishes when taking expectation, yielding

$$E\big(Wf(W)\big) = E\sum_{i=1}^{n} \xi_i^2 \int_0^1 f'\big(W^{(i)} + u\xi_i\big)\, du.$$

On the other hand, again with reference to the left hand side of (1.6), since $\sigma_1^2, \ldots, \sigma_n^2$ sum to 1, and ξ_i and $W^{(i)}$ are independent, we may write

$$\begin{aligned} Ef'(W) &= E\sum_{i=1}^{n} \sigma_i^2 f'(W) \\ &= E\sum_{i=1}^{n} \sigma_i^2 f'\big(W^{(i)}\big) + E\sum_{i=1}^{n} \sigma_i^2 \big(f'(W) - f'\big(W^{(i)}\big)\big) \\ &= E\sum_{i=1}^{n} \xi_i^2 f'\big(W^{(i)}\big) + E\sum_{i=1}^{n} \sigma_i^2 \big(f'(W) - f'\big(W^{(i)}\big)\big). \end{aligned}$$

Taking the difference we obtain the expectation of the left hand side of (1.6) at W,

$$\begin{aligned} E\big(f'(W) - Wf(W)\big) &= E\sum_{i=1}^{n} \xi_i^2 \int_0^1 \big(f'\big(W^{(i)}\big) - f'\big(W^{(i)} + u\xi_i\big)\big)\, du \\ &\quad + E\sum_{i=1}^{n} \sigma_i^2 \big(f'(W) - f'\big(W^{(i)}\big)\big). \end{aligned} \tag{1.7}$$

When n is large, as $\xi_1, \ldots, \xi_n$ are random variables of comparable size, it now becomes apparent why this expectation is small, no matter the distribution of the summands. Indeed, W and $W^{(i)}$ only differ by the single variable ξ_i, accounting for roughly $1/\sqrt{n}$ of the total variance, so the differences in both terms above are small.

To make the case more convincingly, when f has a bounded second derivative, then for all $u \in [0, 1]$, with $\|g\|$ denoting the supremum norm of a function g, the mean value theorem yields

$$\left|f'\left(W^{(i)}\right) - f'\left(W^{(i)} + u\xi_i\right)\right| \leq |\xi_i| \|f''\|.$$

As this bound applies as well to the second term in (1.7), it being the case $u = 1$, when ξ_i has third moments we obtain

$$\begin{aligned} \left|E\left(f'(W) - Wf(W)\right)\right| &\leq \|f''\| \sum_{i=1}^{n} \left(E\left|\xi_i^3\right| + \sigma_i^2 E|\xi_i|\right) \\ &\leq 2\|f''\| \sum_{i=1}^{n} E\left|\xi_i^3\right|, \end{aligned} \tag{1.8}$$

by Hölder's inequality.

The calculation reveals the need for the understanding of the smoothness relation between the solution f and the given function h. For starters, we see directly from (1.6) that f always has one more degree of smoothness than h, which, naturally, helps. However, as the original question was regarding the evaluation of the difference of expectations $Eh(W) - Eh(Z)$ expressed in terms of h, we see that in order to answer using (1.8) that bounds on quantities such as $\|f''\|$ must be provided in terms of some corresponding bound involving h. It is also worth noting that this illustration, and therefore also the original paper of Stein, contains the germ of several of the couplings which we will develop and apply later on, the present one bearing the most similarity to the analysis of local dependence.

The resemblance between Stein's 'leave one out' approach and the method of Lindeberg (see, for instance, Section 8.6 of Breiman 1986) is worth some exploration. Let $X_1, X_2, \ldots$ be i.i.d. mean zero random variables with variance 1, and for each n let

$$\xi_{i,n} = \frac{X_i}{\sqrt{n}}, \quad i = 1, \ldots, n, \tag{1.9}$$

the elements of a triangular array. The basic idea of Lindeberg is to compare the sum

$$W_n = \xi_{1,n} + \cdots + \xi_{n,n}$$

to the sum

$$Z_n = Z_{1,n} + \cdots + Z_{n,n}$$

of mean zero, i.i.d. normals $Z_{1,n}, \ldots, Z_{n,n}$ with $\mathrm{Var}(Z_n) = 1$. Let h be a twice differentiable bounded function on $\mathbb{R}$ such that h'' is uniformly continuous and

$$M = \sup_{x\in\mathbb{R}} \big|h''(x)\big| < \infty. \tag{1.10}$$

For such an h, the quantity

$$\delta(\epsilon) = \sup_{|x-y|\le\epsilon} \big|h''(x) - h''(y)\big|$$

is bounded over $\epsilon \in \mathbb{R}$ and satisfies $\lim_{\epsilon\downarrow 0} \delta(\epsilon) = 0$.

Write the difference $Eh(W_n) - Eh(Z_n)$ as the telescoping sum

$$Eh(W_n) - Eh(Z_n) = E\sum_{i=1}^{n} h(V_{i,n}) - h(V_{i-1,n}), \tag{1.11}$$

where

$$V_{i,n} = \sum_{j=1}^{i} \xi_{j,n} + \sum_{j=i+1}^{n} Z_{j,n},$$

with the usual convention that an empty sum is zero. In this way, the variables interpolate between $W_n = V_{n,n}$ and $Z_n = V_{0,n}$. Writing

$$U_{i,n} = \sum_{j=1}^{i-1} \xi_{j,n} + \sum_{j=i+1}^{n} Z_{j,n},$$

a Taylor expansion on the summands in (1.11) yields

$$\begin{aligned} h(V_{i,n}) - h(V_{i-1,n}) &= h(U_{i,n} + \xi_{i,n}) - h(U_{i,n} + Z_{i,n}) \\ &= (\xi_{i,n} - Z_{i,n})h'(U_{i,n}) \\ &\quad + \frac{1}{2}\xi_{i,n}^2 h''(U_{i,n} + u\xi_{i,n}) - \frac{1}{2}Z_{i,n}^2 h''(U_{i,n} + vZ_{i,n}), \end{aligned}$$

for some $u, v \in [0, 1]$. Since h' can grow at most linearly the expectation of the first term exists, and, as $\xi_{i,n}$ and $Z_{i,n}$ are independent of $U_{i,n}$, equals zero.

Considering the expectation of the remaining second order terms, write

$$E\xi_{i,n}^2 h''(U_{i,n} + u\xi_{i,n}) = E\big(\xi_{i,n}^2 h''(U_{i,n})\big) + \alpha E\big(\xi_{i,n}^2 \delta\big(|\xi_{i,n}|\big)\big),$$

for some $\alpha \in [-1, 1]$, with a similar equality holding for the expectation of the last term. As $E\xi_{i,n}^2 = EZ_{i,n}^2$, taking the difference of the second order terms, using independence, and that $\xi_{i,n}$ and $Z_{i,n}$ are identically distributed, respectively, for $i = 1, \ldots, n$, yields

$$E\big|h(V_{i,n}) - h(V_{i-1,n})\big| \le \frac{1}{2}\big(E\big(\xi_{1,n}^2 \delta\big(|\xi_{1,n}|\big)\big) + E\big(Z_{1,n}^2 \delta\big(|Z_{1,n}|\big)\big)\big). \tag{1.12}$$

Recalling (1.9), we have

$$E\big(\xi_{1,n}^2 \delta\big(|\xi_{1,n}|\big)\big) = \frac{1}{n} E\big(X_1^2 \delta\big(n^{-1/2}|X_1|\big)\big),$$

with a similar equality holding for the second term of (1.12). Hence, by (1.11), with Z now denoting a standard normal variable, summing yields

$$\big|Eh(W_n) - Eh(Z)\big| \le \frac{1}{2}\big(E\big(X_1^2\delta\big(n^{-1/2}|X_1|\big)\big) + E\big(Z^2\delta\big(n^{-1/2}|Z|\big)\big)\big).$$

By (1.10), $\delta(\epsilon) \le 2M$ for all $\epsilon \in \mathbb{R}$, so $X_1^2\delta(n^{-1/2}|X_1|) \le 2MX_1^2$. As $X_1^2\delta(n^{-1/2}|X_1|) \to 0$ almost surely as $n \to \infty$, the dominated convergence theorem implies the first term above tends to zero. Applying the same reasoning to the second term we obtain

$$\lim_{n\to\infty} \big|Eh(W_n) - Eh(Z)\big| = 0. \tag{1.13}$$

As the class of functions h for which we have obtained $Eh(W_n) \to Eh(Z)$ is rich enough, we have shown $W_n \to_d Z$.

Both the Stein and Lindeberg approaches proceed through calculations that 'leave one out.' However, the Stein approach seems more finely tuned to the target distribution, using the solution of a differential equation tailored to the normal. Moreover, use of the Stein differential equation provides that the functions f being evaluated on the variables of interest have one degree of smoothness over that of the basic test functions h which are used to gauge the distance between W and Z. However, the main practical difference between Stein's method and that of Lindeberg, as far as outcome, is the former's additional benefit of providing a bound on the distance to the target, and not only convergence in distribution; witness the difference between conclusions (1.8) and (1.13). Furthermore, Stein's method allows for a variety of ways in which variables can be handled in the Stein equation, the 'leave one out' approach being just the beginning.

1.4 Outline and Summary

We begin in Chap. 2 by introducing and working with the fundamentals of Stein's method. First we prove the Stein characterization (1.4) for the normal, and develop bounds on the Stein equation (1.6) that will be required throughout our treatment; the multivariate Stein equation for the normal, and its solution by the generator method, is also introduced here.

The 'leave one out' coupling considered in Sect. 1.3 is but one variation on the many ways in which variables close to the one of interest can enter the Stein equation, and is in particular related to some of the couplings we consider later on to handle locally dependent variables. Four additional, and somewhat overlapping, basic methods for handling variables in the Stein equation are introduced in Chap. 2: the K-function approach, the original exchangeable pair method of Stein, and the zero bias and size bias transformations. Illustrations of how these methods allow for various manipulations in the Stein equation are provided, as well as a number of examples, some of which will continue as themes and illustrations for the remainder of the book. The independent case, of course, serves as one important testing ground

throughout. A framework that includes some of our approaches is considered in Sect. 2.4. Some technical calculations for bounds to the Stein equation appear in the Appendix to Chap. 2, as do other such calculations in subsequent chapters.

Chapter 3 focuses on the independent case. The goal is to demonstrate a version of the classical Berry–Esseen theorem using Stein's method. Along the way techniques are developed for obtaining L^1 bounds, and the Lindeberg central limit theorem is shown as well. The Berry–Esseen theorem is first demonstrated for the case where the random variables are bounded. The boundedness condition is then relaxed in two ways, first by concentration inequalities, then by induction. This chapter concludes with a lower bound for the Berry–Esseen inequality. As seen in the chapter dependency diagram that follows, Chaps. 2 and 3 form much of the basis of this work.

Chapter 4 develops a theory for obtaining L^1 bounds using the zero bias coupling, and a main result is obtained which can be applied in non-independent settings. A number of examples are presented for illustration. The case of independence is considered first, with an L^1 Berry–Esseen bound followed by the demonstration of a type of contraction principle satisfied by sums of independent variables which implies, or even in a way explains, normal convergence. Bounds in L^1 are then proved for hierarchical structures, that is, self similar, fractal type objects whose scale at small levels is replicated on the larger. Then, making our first departure from independence we prove L^1 bounds for the projections of random vectors having distribution concentrated on regular convex sets in Euclidean space. Next, illustrating a different coupling, L^1 bounds to the normal for the combinatorial central limit theorem are given. Though the combinatorial central limit theorem contains simple random sampling as a particular case, somewhat better bounds may be obtained by applying specifics in the special case; hence, an L^1 bound is given for the case of simple random sampling alone. Next we present Chatterjee's L^1 theorem for functions of independent random variables, and apply it to the approximation of the distribution of the volume covered by randomly placed spheres in the Euclidean torus. Results are then given for sums of locally dependent random variables, with applications including the number of local maxima on a graph. Chapter 4 concludes with a consideration of a class of smooth functions, contained in the one which may be used to determine the L^1 distance, for which convergence to the normal is at the accelerated rate of $1/n$, subject to a vanishing third moment assumption.

The theme of Chap. 5 is to provide upper bounds in the L^∞, or Kolmogorov distance, that can be applied when certain bounded couplings can be constructed. Various bounds to the normal for a random variable W are formed by constructing an auxiliary random variable, say $\widetilde{W}$, on the same space as W. We have in mind here, in particular, the cases where $\widetilde{W}$ has the same distribution as W, or the zero bias or size bias distribution of W. The resulting bound is often interpretable, sometimes directly, as a distance between W and $\widetilde{W}$, a small bound being a reflection of a small distance. Heuristically, being able to make a close coupling to W, shows, in a sense, that perturbing W has only a weak effect. Being able to make a close coupling shows the dependence making up W is weak, and, as a random variable has an approximate normal distribution when it depends on many small weakly dependent

factors, such a W should be approximately normal. The bounded couplings studied in this chapter, ones where $|W - \widetilde{W}| \leq \delta$ with probability one for some δ, and are often much easier to manage than unbounded ones. Chapter 5 provides results when bounded zero bias, exchangeable pair, or size bias couplings can be constructed. The chapter concludes with the use of smoothing inequalities to obtain distances between W and the normal over general function classes, one special case being the derivation of Kolmogorov distance bounds when bounded size bias couplings exist.

Chapter 6 applies the L^∞ results of Chap. 5 to a number of applications, all of which involve dependence. Dependence can loosely be classified into two types, first, the local type, such as when each variable has a small neighborhood outside of which the remaining variables are independent, and second, dependence with a global nature. Chapter 6 deals mainly with global dependence but begins to also touch upon local dependence, a topic more thoroughly explored in Chap. 9. Regarding global dependence, the analysis of the combinatorial central limit theorem, studied in L^1 in Chap. 4, is continued here with the goal of obtaining L^∞ results. Results for the classical case are given, where the permutation is uniformly chosen over the symmetric group, as well as for the case where the permutation is chosen with distribution constant over some conjugacy class, such as the class of involutions. Two approaches are considered, one using the zero bias coupling and one using induction. Normal approximation bounds for the so called lightbulb process are also given in this chapter, again an example of handling global dependence, this time using the size bias coupling. The anti-voter model is also studied, handled by the exchangeable pair technique, as is the binary expansion of a random integer. Results for the occurrences of patterns in graphs and permutations, an example of local dependence, are handled using the size bias method.

Returning to the independent case, and inspired by use of the continuity correction for the normal approximation of the binomial, in Chap. 7 we consider the approximation of independent sums of integer valued random variables by the discretized normal distribution, in the total variation metric. The main result is shown by obtaining bounds between the zero biased distribution of the sum and the normal, and then treating the coupled zero biased variable as a type of perturbation.

Continuing our consideration of the independent case, in Chap. 8 we derive non-uniform bounds for sums of independent random variables. In particular, by use of non-uniform concentration inequalities and the Bennett–Hoeffding inequality we provide bounds for the absolute difference between the distribution function $F(z)$ of a sum of independent variables and the normal $\Phi(z)$, which may depend on $z \in \mathbb{R}$. Non-uniform bounds serve as a counterpoint to the earlier derived supremum norm bounds that are not allowed to vary with z, and give information on how the quality of the normal approximation varies over $\mathbb{R}$.

In Chap. 9 we consider local dependence using the K-function approach, and obtain both uniform and non-uniform Berry–Esseen bounds. The results are applied to certain scan statistics, and yield a general theorem when the local dependence can be expressed in terms of a dependency graph whose vertices are the underlying variables, and where two non-intersecting subsets of variables are independent anytime there is no edge in the graph connecting a element of one subset with the other.

In Chap. 10 we develop uniform and non-uniform bounds for non-linear functions $T(X_1, \ldots, X_n)$, of independent random variables $X_1, \ldots, X_n$, that can be well approximated by a linear term plus a non-linear remainder. Applications include U-statistics, L-statistics and random sums. Randomized concentration inequalities are established in order to develop the theory necessary to cover these examples.

In previous chapters we have measured the accuracy of approximations using differences between two distributions. For the most part, the resulting measures are sensitive to the variations between distributions in their bulk, that is, measures like the L^1 or L^∞ norm typically compare two distributions in the region where most of their mass is concentrated. In contrast, in Chap. 11, we consider moderate deviations of distributions, and rather than consider a difference, compare the ratio of the distribution function of the variable W of interest to that of the normal. Information on small probabilities in the tail become available in this way. Applications of the results of this chapter include the combinatorial central limit theorem, the anti-voter model, the binary expansion of a random integer, and the Curie–Weiss model.

In Chap. 12 we consider multivariate normal approximation, extending both the size bias and exchangeable pair methods to this setting. In the latter case we show how in some cases the exchangeable pair 'linearity condition' can be achieved by embedding the problem in a higher dimension. Applications of both methods are applied to problems in random graphs.

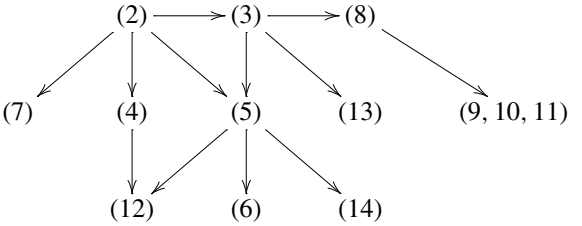
We momentarily depart from normal approximation in Chap. 13. We confine ourselves to approximations by continuous distributions for which the methods of the previous chapters may be extended. As one application, we approximate the distribution of the total spin of the Curie–Weiss model from statistical physics, at the critical inverse temperature, by a distribution with density proportional to $\exp(-x^4/12)$ using the method of exchangeable pairs. We also develop bounds for approximation by the exponential distribution, and apply it to the spectrum of the Bernoulli Laplace Markov chain, and first passage times for Markov chains.

In Chap. 14 we consider two applications of Stein's method, each of which go well beyond the confines of the method's originally intended uses; the approximation of the distribution of characters of elements chosen uniformly from compact Lie groups, and of random variables in a fixed Wiener chaos of Brownian motion, using the tools of Malliavin calculus. Regarding the first topic, the study of random characters is in some sense a generalization to abstract groups of the study of traces of random matrices, a framework into which the combinatorial central limit theorem can be made to fit. As for the second, joining Stein's method to Malliavin calculus shows that the underlying fundamentals of Stein's method, in particular the basic characterization of the normal which can be shown by integration by parts, can be extended, with great benefit, to abstract Wiener spaces.

As for what this book fails to include, narrowing in as it does on what can be shown in the realm of normal approximation by Stein's method, we do not consider, most notably, transform methods, mixing, or martingales. For these topics, having more history than the one presently considered, sources already abound.

We stress to the reader that this book need not at all be read in a linear fashion, especially if one is interested in applications and is willing to forgo the proofs of the

theorems on which the applications are based. The following diagram reflects the dependence of each chapter on the others.



http://www.springer.com/978-3-642-15006-7

Normal Approximation by Stein's Method
Chen, L.H.Y.; Goldstein, L.; Shao, Q.-M.
2011, XII, 408 p. 3 illus., Hardcover
ISBN: 978-3-642-15006-7