ELSEVIER Journal of Computational and Applied Mathematics 55 (1994) 135–143

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

On normal approximation rates for certain sums of dependent random variables☆

Yosef Rinott

Department of Mathematics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0112, United States

Received 1 March 1992; revised 15 January 1993

Abstract

Let $X_1, \ldots, X_n$ be dependent random variables, and set $\lambda = E\{\sum_{i=1}^n X_i\}$, and $\sigma^2 = \mathrm{Var}\{\sum_{i=1}^n X_i\}$. In most of the applications of Stein's method for normal approximations, the error rate $|P((\sum_{i=1}^n X_i - \lambda)/\sigma \leqslant w) - \Phi(w)|$ is of the order of $\sigma^{-1/2}$. This rate was improved by Stein (1986) and others in some special cases. In this paper it is shown that for certain bounded random variables, a simple refinement of error-term calculations in Stein's method leads to improved rates.

Keywords: Stein's method; Dependency graph; Central limit theorem

1. Introduction

Let $X_1, \ldots, X_n$ be dependent random variables, $W = \sum_{i=1}^n X_i$, $EW = \lambda$, and assume $\mathrm{Var}\, W = \sigma^2$ is of the order n. When Stein's method for normal approximation applies, it leads to $|Eh((W-\lambda)/\sigma) - \int h\,\mathrm{d}\Phi| \leqslant Cn^{-1/2}$, for any h having a bounded derivative. However, in many applications in statistics or combinatorics one would like to bound the expression $|P((W-\lambda)/\sigma \leqslant w) - \Phi(w)|$. This requires an approximation which in Stein's formulations often leads to $|P((W-\lambda)/\sigma \leqslant w) - \Phi(w)| \leqslant Cn^{-1/4}$, and clearly, the rate $n^{-1/4}$ (or $\sigma^{-1/2}$) does not appear optimal. In certain cases it is possible to refine the calculation of the convergence rate and obtain the rate of $n^{-1/2}$ (σ^{-1}), or the rate $n^{-(1/2-\delta)}$ for a small $\delta > 0$ in other cases. This paper demonstrates this refinement, which is achieved at the expense of assuming that the random variables are bounded. The present approach can also be attempted under assumptions on moments. We note that in many applications the X_i's are indicators of certain events so they are

☆ The paper has been presented at the Special Session "Applied Probability" of the American Mathematical Society Meeting, University of California at Santa Barbara, November 9–10, 1991, organized by A.P. Godbole and S.T. Rachev.

SSDI 0377-0427(93)E0181-K

obviously bounded. See, for example, [1–3], and references therein, and the example below. We first quote the following result, which is among the most useful formulations of Stein's Central Limit Theorems for dependent random variables.

Theorem 1.1 (Stein [6, p. 110]). *Let $X_1, \ldots, X_n$ be random variables, and let S_i be subsets of $\{1, \ldots, n\}$ such that*

$$EX_i = 0, \qquad EX_i^4 < \infty, \quad i = 1, \ldots, n, \qquad E \sum_{i=1}^{n} X_i \sum_{j \in S_i} X_j = 1.$$

Let $W = \sum_{i=1}^{n} X_i$. Then for any $w \in \mathbb{R}$,

$$|P(W \leqslant w) - \Phi(w)| \leqslant 2\sqrt{E\{\textstyle\sum_{i=1}^{n} \sum_{j \in S_i} (X_i X_j - EX_i X_j)\}^2}$$
$$+ \sqrt{\pi/2}E \sum_{i=1}^{n} |E[X_i \mid X_j\colon j \notin S_i]| + 2^{3/4}\pi^{-1/4}\sqrt{E \textstyle\sum_{i=1}^{n} |X_i| (\sum_{j \in S_i} X_j)^2}. \tag{1}$$

In many applications we have X_i independent of $\{X_j : j \notin S_i\}$, so that S_i can be viewed as a "dependence neighborhood". In this case the second term on the right-hand side of (1) vanishes. If in addition $|S_i| \leqslant m$ where m is some constant, we say that the X_i's exhibit m-dependence. Theorems of the above kind have been successfully applied to combinatorial constructions (e.g., various statistics related to graphs) where $|S_i|$ are slowly growing with n, and where we do not have exchangeability or some other simple structure, and in cases where abstract mixing conditions are hard to study, and do not appear natural.

In order to understand the convergence rate implied by Theorem 1.1, let us consider the case that $Y_1, \ldots, Y_n$ are i.i.d. random variables, $EY_i = 0$, $EY_i^2 = 1$, and $\operatorname{Var} Y_i^4 < \infty$. Then Theorem 1.1 applies with $X_i = Y_i/\sqrt{n}$, and $S_i = \{i\}$. The first term on the right-hand side of (1) is now easily seen to be of order $n^{-1/2}$. The second term vanishes by the independence of the X_i's. However, the last term on the right-hand side of (1) equals $2^{3/4}\pi^{-1/4}\sqrt{E|Y_i|^3}\,n^{-1/4}$. Thus (1) yields the rate $n^{-1/4}$. However, it is well known that in this i.i.d. case the rate should scale like $n^{-1/2}$. It can be shown (see example below) that a similar problem may arise in the case of dependent variables. Stein [6] obtained the rate of $n^{-1/2}$ using his method for i.i.d. random variables; however, the main interest in Stein's method is in the case of dependent variables. The rate $n^{-1/2}$ was obtained in [5] for indicator random variables (which are obviously bounded) under very special dependence conditions. Complex abstract conditions for this rate are discussed in [4].

In Section 2 we present and prove versions of Theorem 1.1, which for bounded random variables may lead to better approximation rates. A similar approach (under suitable conditions) serves to sharpen other results in [2, 6], but this will be done elsewhere.

2. Results and proofs

In certain applications involving bounded dependent (as well as i.i.d.) random variables (see examples below) the following result will produce the desirable $n^{-1/2}$ approximation rate.

Note that the constants B, D, and σ defined in the following two theorems are allowed to depend on n.

Theorem 2.1. *Let $Y_1, \ldots, Y_n$ be random variables such that*

$$EY_i = 0, \quad |Y_i| \leqslant B \text{ a.s.}, \; i = 1, \ldots, n.$$

Let S_i be subsets of $\{1, \ldots, n\}$ and set

$$D = \max_{1 \leqslant i \leqslant n} |S_i|, \qquad \sigma^2 = E \sum_{i=1}^{n} Y_i \sum_{j \in S_i} Y_j, \qquad X_i = Y_i/\sigma, \quad i = 1, \ldots, n, \quad \text{and} \quad W = \sum_{i=1}^{n} X_i.$$

Then for any $w \in \mathbb{R}$,

$$\begin{aligned}
&|P(W \leqslant w) - \Phi(w)| \\
&\leqslant 4\sqrt{E\{\textstyle\sum_{i=1}^n \sum_{j \in S_i} (X_i X_j - E X_i X_j)\}^2} + \sqrt{2\pi} E \sum_{i=1}^{n} |E[X_i \mid X_j \colon j \notin S_i]| \\
&\quad + 2\sqrt{E[\textstyle\sum_{i=1}^n |X_i| (\sum_{j \in S_i} X_j)^2]^2} \{\sqrt{1 + E \textstyle\sum_{i=1}^n \sum_{j \notin S_i} X_j E[X_i \mid X_j \colon j \notin S_i]} + 5\sqrt{2\pi}/4\} \\
&\quad + \frac{1}{\sqrt{2\pi}} \frac{DB}{\sigma} + \frac{4}{\sqrt{2\pi}} \frac{n}{\sigma^3} D^2 B^3.
\end{aligned} \tag{2}$$

The bound in Theorem 2.1 is so written for comparison with that in Theorem 1.1. It is easy to verify that for bounded independent random variables the bound of Theorem 2.1 has the correct order of $n^{-1/2}$. For applications involving dependent X_i's, where we have X_i independent of $\{X_j \colon j \notin S_i\}$, the following formulation is very convenient. We need the following definition.

Definition. Let $\{X_i \colon i \in \mathscr{V}\}$ be a collection of random variables. The graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where $\mathscr{V}$ and $\mathscr{E}$ denote the vertex set and the edge set, respectively, is said to be a *dependency graph* for the collection if for any pair of disjoint subsets of $\mathscr{V}$, A_1 and A_2 such that no edge in $\mathscr{E}$ has one endpoint in A_1 and the other in A_2, the sets of random variables $\{X_i \colon i \in A_1\}$ and $\{X_i \colon i \in A_2\}$ are independent.

The degree $d(v)$ of a vertex v in a graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ is the number of edges connected to this vertex. The maximal degree of a graph is $\max_{v \in \mathscr{V}} d(v)$.

Theorem 2.2. *Let $Y_1, \ldots, Y_n$ be random variables having a dependency graph whose maximal degree is strictly less than D, satisfying $|Y_i - EY_i| \leqslant B$ a.s., $i = 1, \ldots, n$, $E\sum_i^n Y_i = \lambda$ and* $\operatorname{Var} \sum_i^n Y_i = \sigma^2 > 0$. *Then*

$$\left| P\left(\frac{\sum_i^n Y_i - \lambda}{\sigma} \leqslant w \right) - \Phi(w) \right| \leqslant \frac{1}{\sigma} \left\{ \sqrt{\frac{1}{2\pi}} DB + 16 \left(\frac{n}{\sigma^2} \right)^{1/2} D^{3/2} B^2 + 10 \left(\frac{n}{\sigma^2} \right) D^2 B^3 \right\}. \tag{3}$$

Note that when D and B are bounded or are negligible compared to σ, and σ^2 is of the order of n, (3) yields a rate of $1/\sigma$, or equivalently $n^{-1/2}$. In particular we obtain this rate in the case of uniformly bounded independent random variables.

Theorem 2.2 resembles Corollary 2 in [1]. Stein's Theorem 1.1 was used directly to derive Corollary 2 in [1], and yielded an approximation rate which is essentially the square root of the rate obtained now in Theorem 2.2.

Example. The number of local maxima in a graph whose vertices are randomly ranked was studied in [1,2], where the approximation rate of $\sigma^{-1/2}$ was obtained. As an illustration of the possible improvement consider the following simple application of Theorem 2.2. See [1,2] for further details. Assign a random ranking to the vertices of the hypercube $\{0,1\}^N$, set $n = 2^N$, and let Y_i, $i = 1, \ldots, n$, denote the indicator of the event that the ith vertex of the hypercube is a local maximum, that is, its ranking is higher than that of the neighboring vertices in the hypercube. Thus, $\sum_{i=1}^n Y_i$ counts the number of local maxima on the graph. It can be shown that $\sigma^2 = \operatorname{Var} \sum_{i=1}^n Y_i = 2^{N-1}(N-1)/(N+1)^2$. Also, it is not hard to see that Y_i depends only on Y_j's belonging to vertices j in $\{0,1\}^N$ which differ from the vertex i by at most two coordinates. Therefore in Theorem 2.2 we have $D = N + \binom{N}{2} + 1 \leqslant N^2$. Clearly $B = 1$, and from Theorem 2.2 we obtain a normal approximation rate to the standardized number of local maxima which is the order of $N^{5.5}2^{-N/2}$ or equivalently $(\log n)^{5.5} n^{-1/2}$. The square root of this order was obtained in [1,2].

Before deriving Theorem 2.1 we show that it implies Theorem 2.2.

Proof of Theorem 2.2. Let $\mathscr{V} = \{1, \ldots, n\}$ and let $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ be a dependency graph for the Y_i's, having a maximal degree which is strictly less than D. Let $S_i = \{i\} \cup \{j \in \mathscr{V}\colon \{i,j\} \in \mathscr{E}\}$, that is, S_i consists of i and all the vertices connected to i in $\mathscr{G}$. The assumption on the maximal degree of $\mathscr{G}$ implies $|S_i| \leqslant D$ for all i. We can assume, without loss of generality, that $EY_i = 0$, $i = 1, \ldots, n$. Note also that for each i, Y_i is independent of $\{Y_j\colon j \notin S_i\}$ implying $\sigma^2 = E\sum_{i=1}^n Y_i \sum_{j \in S_i} Y_j$. Setting $X_i = Y_i/\sigma$, $i = 1, \ldots, n$, we can now apply Theorem 2.1.

Note that in (2) the two terms involving $E[X_i \mid X_j\colon j \notin S_i]$ will now vanish.

To simplify the first term in (2) set $U_{i,j} = X_i X_j - EX_i X_j$. Then

$$E\left\{\sum_{i=1}^n \sum_{j \in S_i} (X_i X_j - EX_i X_j)\right\}^2 = E \sum_{i=1}^n \sum_{j \in S_i} U_{i,j} \sum_{k=1}^n \sum_{l \in S_k} U_{k,l}. \tag{4}$$

Since $EU_{i,j} = 0$, we have $EU_{i,j}U_{k,l} = 0$ provided $k, l \notin S_i \cup S_j$, by independence. The number of nonzero terms in (4) can be bounded by $4nD^3$ (choose i in n ways, then choose $j \in S_i$ in (at most) D ways, now choose k (or l) in $S_i \cup S_j$ in $2D$ ways and finally l (or k) in S_k (or S_l) in D ways) and each term by $[2(B/\sigma)^2]^2$. Thus the first term in (2) is bounded by $4\sqrt{16(n/\sigma^4)D^3B^4}$.

Next note that $\sqrt{E[\sum_{i=1}^n |X_i| (\sum_{j \in S_i} X_j)^2]^2} \leqslant \sqrt{[n(B/\sigma)(DB/\sigma)^2]^2} = (n/\sigma^3)D^2B^3$. These bounds and some numerical simplifications lead to (3). □

Proof of Theorem 2.1. Most of the calculation follows [6, pp. 105–110]; the point of departure from Stein's calculation will be highlighted by a remark following (11) below. The first step, coupling, involves the introduction of new random variables. Let the random index I be uniformly distributed over $\{1, \ldots, n\}$, independent of the X_i's, and set $W = \sum_{i=1}^n X_i$, $W^* = W - \sum_{j \in S_I} X_j$, and $G = nX_I$.

The relation

$$EWf(W) = EGf(W) \tag{5}$$

for any function f for which the expectations exist, is easy to verify. Let h be piecewise continuously differentiable, and define

$$Nh = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(x) e^{-(1/2)x^2} dx, \qquad f(t) = (U_N h)(t) = e^{(1/2)t^2} \int_{-\infty}^{t} [h(x) - Nh] e^{-(1/2)x^2} dx.$$

The fact that f solves the differential equation $f'(w) - wf(w) = h(w) - Nh$ and (5) imply

$$Eh(W) - Nh = E\{ f'(W) - G[f(W) - f(W^*)] - Gf(W^*)\}.$$

A Taylor series expansion of $f(W) - f(W^*)$ (with integral remainder) yields

$$Eh(W) - Nh = E\{ f'(W)[1 - G(W - W^*)]\} + E\int_{W^*}^{W} G(t - W^*) df'(t) - EGf(W^*). \tag{6}$$

Let $\mathscr{B}$ denote the σ-field generated by the X_i's. Note that $E[G(W - W^*) | \mathscr{B}] = \sum_{i=1}^{n} \sum_{j \in S_i} X_i X_j$ and observe that the assumptions imply $\sum_{i=1}^{n} \sum_{j \in S_i} EX_i X_j = 1$. We shall now assume that h is nonnegative and $\sup h \leqslant 1$. With the bound $\sup |(U_N h)'| \leqslant 2 \sup |h - Nh|$ [6, p. 25], applied to such an h, we obtain

$$\begin{aligned} &|E\{ f'(W)[1 - G(W - W^*)]\}| \\ &\quad \leqslant 2\sqrt{E\{1 - E[G(W - W^*) | \mathscr{B}]\}^2} = 2\sqrt{E\{\textstyle\sum_{i=1}^{n} \sum_{j \in S_i} (X_i X_j - EX_i X_j)\}^2}. \end{aligned} \tag{7}$$

In order to bound the last term in (6), recall the inequality $\sup |U_N h| \leqslant \sqrt{\pi/2} \sup |h - Nh|$ [6, p. 25], which implies

$$\begin{aligned} |EGf(W^*)| &= |E \textstyle\sum_{i=1}^{n} f(\sum_{j \notin S_i} X_j) E[X_i | X_j: j \notin S_i]| \\ &\leqslant \sqrt{\pi/2} E \sum_{i=1}^{n} |E[X_i | X_j: j \notin S_i]|. \end{aligned} \tag{8}$$

We now apply (6)–(8) with the function

$$h(x) = \begin{cases} 1 & \text{if } x \leqslant w, \\ 1 - (1/\varepsilon)(x - w) & \text{if } w \leqslant x \leqslant w + \varepsilon, \\ 0 & \text{otherwise.} \end{cases}$$

Since $Nh \leqslant \Phi(w) + \varepsilon/2\sqrt{2\pi}$, and $P(W \leqslant w) \leqslant Eh(W)$, we obtain

$$\begin{aligned} P(W \leqslant w) - \Phi(w) \leqslant{}& \frac{\varepsilon}{2\sqrt{2\pi}} + 2\sqrt{E\{\textstyle\sum_{i=1}^{n} \sum_{j \in S_i} (X_i X_j - EX_i X_j)\}^2} \\ &+ \sqrt{\pi/2} E \sum_{i=1}^{n} |E[X_i | X_j: j \notin S_i]| + E\int_{W^*}^{W} G(t - W^*) df'(t). \end{aligned} \tag{9}$$

We now deviate from Stein's calculations of the term $E\int_{W^*}^{W} G(t - W^*)\,\mathrm{d}f'(t)$. First observe that the relation $f'(w) - wf(w) = h(w) - Nh$ implies

$$f''(t) = -\frac{1}{\varepsilon}\psi(t) + (U_N h)(t) + t(U_N h)'(t), \tag{10}$$

where ψ denotes the indicator function of the interval $[w, w+\varepsilon]$, and therefore

$$E\int_{W^*}^{W} G(t - W^*)\,\mathrm{d}f'(t) = E\int_{W^*}^{W} nX_I(t - W^*)\left\{-\frac{1}{\varepsilon}\psi(t) + (U_N h)(t) + t(U_N h)'(t)\right\}\mathrm{d}t. \tag{11}$$

Remark. It is essentially the $1/\varepsilon$ in the first term which caused the loss in the rate in Stein's calculations. In the calculation of this term below, the refinement consists of taking account of the fact that $\int_{W^*}^{W} nX_I(t - W^*)(1/\varepsilon)\psi(t)$ vanishes unless $\{(W \wedge W^*, W \vee W^*)\cap(w, w+\varepsilon) \neq \emptyset\}$ and the latter event has a small probability.

We first treat the second and third terms in (11). We bound the second term in (11) by using again $\sup|U_N h| \leqslant \sqrt{\pi/2}\sup|h - Nh| \leqslant \sqrt{\pi/2}$ [6, p. 25] to obtain

$$\left|E\int_{W^*}^{W} nX_I(t - W^*)(U_N h)(t)\,\mathrm{d}t\right| \leqslant \sqrt{\pi/2}\,nE|X_I|\tfrac{1}{2}(W - W^*)^2 = \tfrac{1}{2}\sqrt{\pi/2}\,E\sum_{i=1}^{n}|X_i|\Big(\sum_{j\in S_i} X_j\Big)^2. \tag{12}$$

Next, we discuss the third term in (11). We shall attempt only a simple, crude bound here. Simple manipulations and integration by parts of the term $\int_{W^*}^{W}(t - W^*)(t - W)(U_N h)'(t)\,\mathrm{d}t$ below yield

$$\begin{aligned}
&\int_{W^*}^{W}(t - W^*)t(U_N h)'(t)\,\mathrm{d}t\\
&\quad= \int_{W^*}^{W}(t - W^*)(t - W + W)(U_N h)'(t)\,\mathrm{d}t\\
&\quad= \int_{W^*}^{W}(t - W^*)W(U_N h)'(t)\,\mathrm{d}t + \int_{W^*}^{W}(t - W^*)(t - W)(U_N h)'(t)\,\mathrm{d}t\\
&\quad= \int_{W^*}^{W}(t - W^*)W(U_N h)'(t)\,\mathrm{d}t - \int_{W^*}^{W}(t - W^*)(U_N h)(t)\,\mathrm{d}t - \int_{W^*}^{W}(t - W)(U_N h)(t)\,\mathrm{d}t\\
&\quad= \int_{W^*}^{W}(t - W^*)W(U_N h)'(t)\,\mathrm{d}t - 2\int_{W^*}^{W}(t - W^*)(U_N h)(t)\,\mathrm{d}t + \int_{W^*}^{W}(W - W^*)(U_N h)(t)\,\mathrm{d}t.
\end{aligned} \tag{13}$$

Replacing $|U_n h|$ and $|U_N h)'|$ by their respective bounds $\sqrt{\frac{1}{2}\pi}$ and 2, integrating, and combining the last two terms, we have

$$\begin{aligned}
&\left|E\int_{W^*}^{W} nX_I(t-W^*)t(U_N h)'(t)\,\mathrm{d}t\right| \\
&\quad\leqslant nE|X_I(W-W^*)^2W| + 2\sqrt{\pi/2}\,nE|X_I|(W-W^*)^2 \\
&\quad\leqslant \sqrt{EW^2E[EnX_I(W-W^*)^2\,|\,\mathscr{B}]^2} + \sqrt{2\pi}\,nE|X_I|(W-W^*)^2 \\
&\quad= \sqrt{EW^2E[\textstyle\sum_{i=1}^n |X_i|(\sum_{j\in S_i}X_j)^2]^2} + \sqrt{2\pi}\,E\displaystyle\sum_{i=1}^n |X_i|\Big(\sum_{j\in S_i}X_j\Big)^2.
\end{aligned}\tag{14}$$

In (14) the second inequality follows from the Cauchy–Schwarz inequality. (A tighter but more cumbersome bound, involving only fourth-order moments could be calculated here.) Note that

$$EW^2 = E\sum_{i=1}^n\sum_{j\in S_i}X_iX_j + E\sum_{i=1}^n\sum_{j\notin S_i}X_iX_j = 1 + E\sum_{i=1}^n\sum_{j\notin S_i}X_jE[X_i\,|\,X_j: j\notin S_i].\tag{15}$$

Returning to (11) we now study the term $(1/\varepsilon)E\int_{W^*}^{W} nX_I(t-W^*)\psi(t)\,\mathrm{d}t$. It can be written as $(1/\varepsilon)E\int_{W^*}^{W} nX_I(t-W^*)\psi(t)\,\mathrm{d}t\,\mathscr{I}_Q$, where $Q=\{(W\wedge W^*, W\vee W^*)\cap(w,w+\varepsilon)\neq\emptyset\}$ and $\mathscr{I}_Q$ denotes the indicator of the event Q. Observe that $Q=\{W^*\wedge W\leqslant w+\varepsilon\}\cap\{W^*\vee W\geqslant w\}$. Also, setting $U=W-W^*$ we have $W-|U|\leqslant W^*\wedge W$, $W+|U|\geqslant W^*\vee W$. It is now clear that $Q\subseteq\{|W-w|\leqslant|U|+\varepsilon\}=R$, say, and integration yields

$$\left|\frac{1}{\varepsilon}E\int_{W^*}^{W} nX_I(t-W^*)\psi(t)\,\mathrm{d}t\right| \leqslant \frac{n}{2\varepsilon}E|X_I|(W-W^*)^2\mathscr{I}_R.\tag{16}$$

At this point we invoke the bounds $X_i\leqslant B/\sigma$ and $|S_i|\leqslant D$, and deduce that the right-hand side of (16) is bounded by $(n/2\varepsilon\sigma^3)D^2B^3P(R)$. Set

$$\Delta = \sup_w|P(W\leqslant w)-\Phi(w)|.$$

Since $|U|\leqslant DB/\sigma$ and the standard normal density is bounded by $1/\sqrt{2\pi}$, it is easy to see that $P(R)\leqslant\Phi(w+DB/\sigma+\varepsilon)-\Phi(w-DB/\sigma-\varepsilon)+2\Delta\leqslant\sqrt{2/\pi}\,DB/\sigma+\sqrt{2/\pi}\,\varepsilon+2\Delta$.

Combining the results from (9) on, and for simplicity using the bound $E\sum_{i=1}^n|X_i|(\sum_{j\in S_i}X_j)^2\leqslant\sqrt{E[\sum_{i=1}^n|X_i|(\sum_{j\in S_i}X_j)^2]^2}$ we obtain

$$\begin{aligned}
&P(W\leqslant w)-\Phi(w) \\
&\quad\leqslant \frac{\varepsilon}{2\sqrt{2\pi}} + 2\sqrt{E\{\textstyle\sum_{i=1}^n\sum_{j\in S_i}(X_iX_j-EX_iX_j)\}^2} + \sqrt{\pi/2}\,E\displaystyle\sum_{i=1}^n|E[X_i\,|\,X_j: j\notin S_i]| \\
&\qquad+\sqrt{E[\textstyle\sum_{i=1}^n|X_i|(\sum_{j\in S_i}X_j)^2]^2}\,\{\sqrt{1+E\sum_{i=1}^n\sum_{j\notin S_i}X_jE[X_i\,|\,X_j: j\notin S_i]}+5\sqrt{2\pi}/4\} \\
&\qquad+\frac{n}{2\varepsilon\sigma^3}D^2B^3(\sqrt{2/\pi}\,DB/\sigma+\sqrt{2/\pi}\,\varepsilon+2\Delta).
\end{aligned}\tag{17}$$

The above calculation applied also to the function

$$h(x) = \begin{cases} 1 & \text{if } x \leqslant w - \varepsilon, \\ 1 - (1/\varepsilon)(x - w + \varepsilon) & \text{if } w - \varepsilon \leqslant x \leqslant w, \\ 0 & \text{otherwise,} \end{cases}$$

to obtain a lower bound analogous to (17), shows that Δ is bounded by the right-hand side of (17) and we obtain

$$\begin{aligned} &\Delta\left(1 - \frac{n}{\varepsilon\sigma^3} D^2 B^3\right) \\ &\leqslant \frac{\varepsilon}{2\sqrt{2\pi}} + 2\sqrt{E\{\textstyle\sum_{i=1}^n \sum_{j\in S_i}(X_iX_j - EX_iX_j)\}^2} + \sqrt{\pi/2}E \sum_{i=1}^n |E[X_i \mid X_j: j \notin S_i]| \\ &\quad + \sqrt{E[\textstyle\sum_{i=1}^n |X_i|(\sum_{j\in S_i} X_j)^2]}\{\sqrt{1 + E\textstyle\sum_{i=1}^n \sum_{j\notin S_i} X_j E[X_i \mid X_j: j\notin S_i]} + 5\sqrt{2\pi}/4\} \\ &\quad + \sqrt{\frac{1}{2\pi}\frac{n}{\varepsilon\sigma^4}}D^3B^4 + \sqrt{\frac{1}{2\pi}\frac{n}{\sigma^3}}D^2B^3. \end{aligned} \tag{18}$$

Choosing $\varepsilon = 2(n/\sigma^3)D^2B^3$ (which is not optimal, but close enough and simple), we obtain (2) by straightforward calculations. □

Remark. Suppose σ^2 is of the order of n. Note that in the derivation of (17) we used the boundedness of the random variables to obtain $P(R) \leqslant \sqrt{2/\pi}\,DB/\sigma + \sqrt{2/\pi}\varepsilon + 2\Delta$. If one simply uses $P(R) \leqslant 1$ (not taking advantage of the smallness of $P(R)$) and chooses $\varepsilon = \sigma^{-1/2}$, one immediately obtains the rate $\sigma^{-1/2}$, or equivalently $n^{-1/4}$. As explained in the previous remark this is essentially the way this rate was obtained in [6]. For independent random variables $P(R)$ can be appropriately bounded without assuming boundedness. Since the independent case is well known we shall comment on this very briefly. In the bounded case, the event R can essentially be thought of as being $R = \{|W| \leqslant |U|\}$ where for the some constant C we have $|U| \leqslant C$, and the density of $|W|$ is bounded (by 1, say). Then $P(R) \leqslant P(|W| \leqslant C) \leqslant C$ provides the needed bound. In the case of independent unbounded random variables easy manipulations reduce the calculations of $P(R)$ to the case that W and (a modified) U are independent, $|W|$ has a bounded density as before, and $|U|$ is unbounded now, but $E|U| < \infty$. In this case the desired result will follow from $P(R) = EP(|W| \leqslant |U| \mid U) \leqslant E|U|$.

Acknowledgements

I wish to thank Patrick Fitzsimmons whose suggestions simplified and clarified the proofs. This work was supported in part by NSF grant DMS-9001274.

References

[1] P. Baldi and Y. Rinott, On normal approximations of distributions in terms of dependency graphs, *Ann. Probab.* **17** (1989) 1646–1650.
[2] P. Baldi, Y. Rinott and C. Stein, A normal approximation for the number of local maxima of a random function on a graph, in: T.W. Anderson, K.B. Athreya and D.L. Iglehart, Eds., *Probability, Statistics and Mathematics, Papers in Honor of Samuel Karlin* (Academic Press, New York, 1989) 59–81.
[3] A.D. Barbour, M. Karoński and A. Ruciński, A central limit theorem for decomposable random variables with applications to random graphs, *J. Combin. Theory Ser. B* **47** (1989) 125–145.
[4] L.H.Y. Chen, The rate of convergence in the central limit theorem for dependent random variables with arbitrary index set, preprint, 1991.
[5] W. Kordecki, Normal approximation and isolated vertices in random graphs, in: M. Karoński, J. Jaworski and A. Ruciński, Eds., *Random Graphs 1987* (Wiley, New York, 1990) 131–139.
[6] C. Stein, *Approximate Computation of Expectations* (IMS, Hayward, CA, 1986).