

Local limit theorems via Landau–Kolmogorov inequalities

ADRIAN RÖLLIN¹ and NATHAN ROSS²

¹*Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117546. E-mail: adrian.roellin@nus.edu.sg; url: www.stat.nus.edu.sg/~staar*

²*Department of Mathematics and Statistics, University of Melbourne, VIC 3010, Australia. E-mail: nathan.ross@unimelb.edu.au; url: www.ms.unimelb.edu.au/~rossn1*

In this article, we prove new inequalities between some common probability metrics. Using these inequalities, we obtain novel local limit theorems for the magnetization in the Curie–Weiss model at high temperature, the number of triangles and isolated vertices in Erdős–Rényi random graphs, as well as the independence number in a geometric random graph. We also give upper bounds on the rates of convergence for these local limit theorems and also for some other probability metrics. Our proofs are based on the Landau–Kolmogorov inequalities and new smoothing techniques.

Keywords: Curie–Weiss model; Erdős–Rényi random graph; Kolmogorov metric; Landau–Kolmogorov inequalities; local limit metric; total variation metric; Wasserstein metric

1. Introduction

If two probability distributions are close in some metric are they also close in other stronger or different metrics? General inequalities between many common probability metrics are known; see for example Gibbs and Su [17] for a compilation of such results. But one may wonder if it is possible to sharpen such inequalities by imposing simple conditions on the distributions under consideration. An early attempt in this direction was made by McDonald [24], who was able to deduce a local limit theorem for sums of integer valued random variables from a central limit theorem by imposing an additional “smoothness” condition on the distribution of the sum. In this article, we take this approach much further by providing general inequalities between some common probability metrics with integer support that contain an additional factor that measures the “smoothness” of the distributions under consideration; the smaller this factor is, the better the bounds obtained.

To state a simple version of our main result, we need some basic notation. For a function f with domain the integers, denote for $1 \leq p < \infty$,

$$\|f\|_p = \left(\sum_{i \in \mathbb{Z}} |f(i)|^p \right)^{1/p}$$

and $\|f\|_\infty = \sup_{i \in \mathbb{Z}} |f(i)|$, and also define the operators Δ^n recursively by

$$\Delta^0 f(k) = f(k) \quad \text{and} \quad \Delta^{n+1} f(k) = \Delta^n f(k+1) - \Delta^n f(k).$$

A consequence of our main theoretical result, Theorem 2.2 below, is that if F and G are distribution functions of integer supported distributions, then for some universal constant C ,

$$\|\Delta F - \Delta G\|_\infty \leq C \|F - G\|_\infty^{1/2} (\|\Delta^3 F\|_1 + \|\Delta^3 G\|_1)^{1/2}. \tag{1.1}$$

Here $(\|\Delta^3 F\|_1 + \|\Delta^3 G\|_1)^{1/2}$ is the smoothing factor referred to above, so the inequality says that if we can bound it and the supremum of the pointwise differences of the distribution functions F and G (called the *uniform* or *Kolmogorov metric*), then we have a bound on the left-hand side of (1.1), the supremum of the differences of point probabilities; the latter is a quantity that will allow us to obtain local limit theorems.

In practice, it may appear difficult to obtain bounds on the smoothing term since it is defined in terms of quantities we wish to study. In this article, we think of F as being a complicated distribution of interest (e.g., the number of triangles in a random graph model) and of G as a well-known distribution which we are using to approximate F (e.g., a discretized normal or a translated Poisson distribution). Thus, bounding $\|\Delta^3 G\|_1$ should not be difficult – we provide what is needed for our theory and applications in Lemma 4.1 below – so the only real difficulty in using (1.1) in application is bounding $\|\Delta^3 F\|_1$ and in Section 3 we develop tools for this purpose.

To get a sense of the style of result we aim to achieve, we apply (1.1) in the setting of the approximation of the binomial distribution by the normal, where much is known.

Binomial local limit theorem

Let $X \sim \text{Bi}(n, p)$ and let Y have a discretized normal distribution with mean $\mu := np$ and variance $\sigma^2 := np(1 - p)$, that is,

$$\mathbb{P}(Y = k) = \frac{1}{\sqrt{2\pi}} \int_{(k-1/2-\mu)/\sigma}^{(k+1/2-\mu)/\sigma} e^{-u^2/2} du. \tag{1.2}$$

If F and G are the distribution functions of X and Y , then it is well known that $\|F - G\|_\infty \asymp \sigma^{-1} \asymp n^{-1/2}$; here and below the limits and asymptotics are as $n \rightarrow \infty$. Also $\Delta^3 G(k) = \Delta^2 \mathbb{P}(Y = k)$ and some basic calculus and (1.2) imply that $\|\Delta^2 \mathbb{P}(Y = \cdot)\|_1 \asymp \sigma^{-2} \asymp n^{-1}$. Due to the closeness of the binomial distribution to the normal, we anticipate $\|\Delta^3 F\|_1$ to be of this same order as $\|\Delta^3 G\|_1$ and in fact Proposition 3.8 below bounds this term as $\|\Delta^3 F\|_1 = O(\sigma^{-2})$. Putting this all into (1.1), we have that

$$\|\Delta F - \Delta G\|_\infty = O(\sigma^{-3/2}) = O(n^{-3/4}).$$

In fact, it is well known that

$$\|\Delta F - \Delta G\|_\infty \asymp \sigma^{-2} \asymp n^{-1}.$$

This example illustrates that we do not expect our approach to yield tight rates in application. However, our purpose here is to provide a method that can be applied to yield new convergence

results where little is known and, as a by-product of our method of proof, to give some upper bounds on the rates of convergence. We emphasize that apart from well known results about sums of independent random variables, rates of convergence in local limit theorems are not common in the literature: such results are typically difficult to obtain. To the best of our knowledge, all of our results and upper bounds on rates are new. Outside of a few remarks we will not address the interesting but more theoretical question of the optimality of the bounds obtained – we shall focus on applications.

The remainder of the paper is organized as follows. In Section 2, we prove our main theoretical results, inequalities of the form (1.1); these will follow from discrete versions of the classical Landau–Kolmogorov inequalities. In Section 3, we develop tools to bound $\|\Delta^3 F\|_1$ and the analogous quantities appearing on the right hand side of generalizations of (1.1). In Section 4, we illustrate our approach in a few applications, in particular we obtain new local limit theorems with bounds on the rates of convergence for the magnetization in the Curie–Weiss model, the number of isolated vertices and triangles in Erdős–Rényi random graphs and the independence number of a geometric random graph. We also obtain other new limit theorems and bounds on rates for some of these applications.

2. Main result

Our main theoretical result is easily derived from a discrete version of the classical Landau inequality (see Hardy, Landau and Littlewood [19], Section 3) which relates the norm of a function with that of its first and second derivatives. There are many extensions and embellishments of this inequality in the analysis literature; see Kwong and Zettl [22] for a book length treatment.

Theorem 2.1 (Kwong and Zettl [22], Theorem 4.1). *Let k and n be integers with $1 \leq k < n$ and let $1 \leq p, q, r \leq \infty$ given. There is a positive number $C := C(n, k, p, q, r)$ such that*

$$\|\Delta^k f\|_q \leq C \|f\|_p^\alpha \|\Delta^n f\|_r^\beta \quad (2.1)$$

for all $f: \mathbb{Z} \rightarrow \mathbb{R}$ with $\|f\|_p < \infty$ and $\|\Delta^n f\|_r < \infty$, if and only if

$$\frac{n}{q} \leq \frac{n-k}{p} + \frac{k}{r},$$

$\alpha = 1 - \beta$ and

$$\beta = \frac{k - 1/q + 1/p}{n - 1/r + 1/p}.$$

Remark 2.1. Much of the literature surrounding these inequalities is concerned with finding the optimal value of the constant C . In the case that $n = 2$ and either $p = q = r = 1$ or $n = 3$, $p = q = \infty$, and $r = 1$, we can take $C = \sqrt{2}$; see Kwong and Zettl [22], Theorem 4.2. These are two of the main cases discussed below. Also, an inductive argument in n implies that in the former case above we may take $C = 2^{(n-1)/2}$ for $n \geq 2$ and in the latter $C = 2^{(n-2)/2}$ for $n \geq 3$.

These facts are not critical in what follows, so for the sake of simplicity we do not discuss such constants in further detail.

The key connection between Theorem 2.1 and what will follow is that if F and G are distribution functions of integer supported probability distributions, then some well-known probability metrics can be expressed as

$$\begin{aligned} d_K(F, G) &= \|F - G\|_\infty && \text{(Kolmogorov metric),} \\ d_W(F, G) &= \|F - G\|_1 && \text{(Wasserstein metric),} \\ d_{loc}(F, G) &= \|\Delta F - \Delta G\|_\infty && \text{(local metric),} \\ d_{TV}(F, G) &= \frac{1}{2} \|\Delta F - \Delta G\|_1 && \text{(total variation metric).} \end{aligned}$$

Note that $d_{loc}(F, G)$ is the supremum of point probabilities between the distributions given by F and G and is the appropriate metric to use to show local limit theorems.

We are now in a position to state our main theoretical result, but first a last bit of notation. Let F be a distribution function with support on \mathbb{Z} . If m is a positive integer, define

$$\bar{F}^m(j) = \frac{\bar{F}(j) + \dots + F(j - m + 1)}{m};$$

this is the distribution function of the convolution of F with the uniform distribution on $\{0, \dots, m - 1\}$. Note that $\bar{F}^1 = F$ and that if the integer valued random variable X has distribution function F , then

$$\Delta \bar{F}^m(j) = \frac{1}{m} \mathbb{P}(j - m + 1 < X \leq j + 1). \tag{2.2}$$

Theorem 2.2. *If $l \geq 1$ and $m \geq 1$ are integers, then there is a constant $C > 0$ such that, for all distribution functions F and G of integer supported probability distributions,*

$$d_1(\bar{F}^m, \bar{G}^m) \leq C d_2(F, G)^{1-\beta} (\|\Delta^{l+1} \bar{F}^m\|_1 + \|\Delta^{l+1} \bar{G}^m\|_1)^\beta$$

for the following combinations of d_1, d_2 and β :

	d_1	d_2	β
(i)	d_{loc}	d_{TV}	$1/l$
(ii)	d_{loc}	d_K	$1/l$
(iii)	d_{loc}	d_W	$2/(l + 1)$
(iv)	d_{TV}	d_W	$1/(l + 1)$
(v)	d_K	d_W	$1/(l + 1)$

(2.3)

Proof. To prove (ii)–(iv), apply Theorem 2.1 to the function $\bar{F}^m - \bar{G}^m$, with $n = l + 1, k = r = 1$ and the following values of p and q :

$$(ii) \quad q = \infty, p = \infty, \quad (iii) \quad q = \infty, p = 1, \quad (iv) \quad q = 1, p = 1;$$

then use $d_2(\bar{F}^m, \bar{G}^m) \leq d_2(F, G)$ and the triangle inequality. For (i) and (v) use (ii) and (iv), respectively, and then use the fact that $d_K \leq d_{TV}$. \square

Remarks.

1. We mainly use Theorem 2.2 with $m = 1$, where its meaning is most transparent. For $m > 1$, the following direct consequence of (2.2) shows that \bar{F}^m can be used to prove local limit theorems for “clumped” probabilities where the corresponding pointwise results may not hold.

Lemma 2.3. *If X and Y are integer valued random variables with respective distribution functions F and G , then*

$$\sup_{k \in \mathbb{Z}} |\mathbb{P}(k < X \leq k + m) - \mathbb{P}(k < Y \leq k + m)| = m d_{\text{loc}}(\bar{F}^m, \bar{G}^m).$$

2. Theorem 2.2 is really a special case of Theorem 2.1 with $k = r = 1$, but it is clear that similar statements hold by applying Theorem 2.1 to other values of k and r . We choose the value $r = 1$ because we are able to bound $\|\Delta^n F\|_1$. Using the obvious inequality $\|\Delta^n F\|_\infty \leq \|\Delta^{n+1} F\|_1$, we could also usefully apply Theorem 2.1 with $r = \infty$, but this change has no effect on the value of β for a given k, q and p . The term $\|\Delta^2 F\|_\infty$ also appears in the local limit theorem results of McDonald [24] and Davis and McDonald [13]. However, the crucial advantage of $\|\Delta^n F\|_1$ over $\|\Delta^n F\|_\infty$ is that the former is – as we will show – amenable to bounds via probabilistic techniques, whereas the latter seems difficult to handle directly.
3. Inequality (2.1) cannot be improved in general, but since we are considering such inequalities only over the class of functions that are the difference of two distribution functions, it is possible that Theorem 2.2 could be sharpened, either in increasing the exponents or decreasing the constants. Also note that using the triangle inequality in Theorem 2.2 causes some loss of sharpness, but we gain the ability to bound the terms appearing in application, which is our main focus.

3. Estimating the measure of smoothness

In this section, we develop techniques to bound $\|\Delta^n \bar{F}^m\|_1$. Our main tools are Theorems 3.6 and 3.7 below but first we state some simple results. To lighten the notation somewhat, write

$$D_{n,m}(F) := m \|\Delta^{n+1} \bar{F}^m\|_1,$$

or for a random variable W with distribution function F , write $D_{n,m}(\mathcal{L}(W))$ for $D_{n,m}(F)$, and D_n for $D_{n,1}$. Furthermore, define recursively the difference operators

$$\Delta_m^n F(j) = \Delta_m^{n-1} F(j + m) - \Delta_m^{n-1} F(j),$$

where $\Delta_m^0 F(j) = F(j)$.

Note that for a random variable W ,

$$D_1(\mathcal{L}(W)) = 2d_{TV}(\mathcal{L}(W), \mathcal{L}(W + 1)).$$

By a well-known representation of the total variation distance (see, for example, Gibbs and Su [17]), we have

$$D_1(\mathcal{L}(W)) = \sup_{\|g\|_\infty \leq 1} \mathbb{E}\Delta g(W). \tag{3.1}$$

Some of our techniques below are extensions of those for bounding the quantity on the right-hand side (3.1), and so we use the following generalization of (3.1) in bounding $D_{n,m}(F)$.

Lemma 3.1. *Let n and m be nonnegative integers. Let W be a random variable with integer support. Then*

$$D_{n,m}(\mathcal{L}(W)) = \sup_{\|g\|_\infty \leq 1} \mathbb{E}\Delta_m^n g(W).$$

Proof. We only show the case $n = 1$; the general case is similar. Denote $p_i = \mathbb{P}[W = i]$ for $i \in \mathbb{Z}$. We have

$$\begin{aligned} \mathbb{E}\Delta_m g(W) &= \sum_{i \in \mathbb{Z}} p_i (g(i + m) - g(i)) \\ &= \sum_{i \in \mathbb{Z}} (p_{i-m} - p_i)g(i) \\ &= m \sum_{i \in \mathbb{Z}} \left(\frac{p_{i-m} + \dots + p_{i-1}}{m} - \frac{p_{i-m+1} + \dots + p_i}{m} \right) g(i) \\ &= m \sum_{i \in \mathbb{Z}} (\Delta \bar{F}^m(i - 2) - \Delta \bar{F}^m(i - 1))g(i) \\ &= -m \sum_{i \in \mathbb{Z}} \Delta^2 \bar{F}^m(i - 2)g(i), \end{aligned}$$

where in the fourth equality we have used (2.2). We see that for all g such that $\|g\|_\infty \leq 1$, $\mathbb{E}\Delta_m g(W) \leq m\|\Delta^2 \bar{F}^m\|_1$, and choosing $g(i) = -\text{sgn} \Delta^2 \bar{F}^m(i - 2)$ implies the claim. \square

The following sequence of lemmas provide tools for bounding $D_{n,m}(F)$. The proofs are mostly straightforward. We assume that all random variables are integer valued.

Lemma 3.2. *Let n and m be positive integers. If W is a random variable, then*

$$D_{n,m}(\mathcal{L}(W)) \leq mD_{n,1}(\mathcal{L}(W)).$$

Proof. If W has distribution function F , then the triangle inequality implies

$$\begin{aligned} D_{n,m}(\mathcal{L}(W)) &= \sum_{k \in \mathbb{Z}} |\Delta^{n+1} F(k) + \dots + \Delta^{n+1} F(k+m-1)| \\ &\leq \sum_{k \in \mathbb{Z}} |\Delta^{n+1} F(k)| + \dots + \sum_{k \in \mathbb{Z}} |\Delta^{n+1} F(k+m-1)| \\ &= m D_{n,1}(F) = m D_{n,1}(\mathcal{L}(W)). \end{aligned} \quad \square$$

Lemma 3.3. *Let n and m be positive integers. If W is a random variable and \mathcal{F} is a σ -algebra, then*

$$D_{n,m}(\mathcal{L}(W)) \leq \mathbb{E} D_{n,m}(\mathcal{L}(W|\mathcal{F})).$$

Proof. If f is a bounded function, then

$$|\mathbb{E} \Delta_m^n f(W)| \leq \mathbb{E} |\mathbb{E} [\Delta_m^n f(W) | \mathcal{F}]| \leq \|f\|_\infty \mathbb{E} D_{n,m}(\mathcal{L}(W|\mathcal{F})).$$

By Lemma 3.1, the claim follows. □

Lemma 3.4. *If X_1 and X_2 are independent random variables, then, for all $n_1, n_2, m \geq 1$,*

$$D_{n_1+n_2,m}(\mathcal{L}(X_1 + X_2)) \leq D_{n_1,m}(\mathcal{L}(X_1)) D_{n_2,m}(\mathcal{L}(X_2)). \tag{3.2}$$

If X_1, \dots, X_N is a sequence of independent random variables and $n \leq N$,

$$D_{n,m}(\mathcal{L}(X_1 + \dots + X_N)) \leq \prod_{i=1}^n D_{1,m}(\mathcal{L}(X_i)). \tag{3.3}$$

Proof. Let f be a bounded function and define

$$g(x) := \mathbb{E} \Delta_m^{n_2} f(x + X_2) = \sum_{j \in \mathbb{Z}} \Delta_m^{n_2} f(x + j) \mathbb{P}(X_2 = j).$$

Note that $\|g\|_\infty \leq D_{n_2,m}(X_2) \|f\|_\infty$ and we claim

$$\mathbb{E} \Delta_m^{n_1+n_2} f(X_1 + X_2) = \mathbb{E} \Delta_m^{n_1} g(X_1),$$

which follows by independence (that is, the conditioning has no effect). Hence,

$$D_{n_1+n_2,m}(\mathcal{L}(X_1 + X_2)) \leq D_{n_1,m}(X_1) \|g\|_\infty \leq D_{n_1,m}(X_1) D_{n_2,m}(X_2) \|f\|_\infty$$

which proves (3.2). A similar argument establishes that $D_{n,m}(X_1 + X_2) \leq D_{n,m}(X_1)$ so now (3.3) follows by induction. □

The quantity $D_1(\mathcal{L}(W), \mathcal{L}(W + 1)) = 2d_{TV}(W, W + 1)$ has appeared in extending the central limit theorem for sums of integer valued random variables to stronger metrics such as the total variation and local limit metric; see, for example, Barbour and Xia [5], Barbour and Čekanavičius [2] and Goldstein and Xia [18]. In these cases, the main tool for bounding $D_1(\mathcal{L}(W))$ was initially the Mineka coupling but the following result is now the best available (see Pósfai [28] and references there).

Lemma 3.5 (Mattner and Roos [23], Corollary 1.6). *Let X_1, X_2, \dots, X_N be a sequence of independent integer valued random variables and $S_N = \sum_{i=1}^N X_i$. Then*

$$D_1(\mathcal{L}(S_N)) = D_{1,1}(\mathcal{L}(S_N)) \leq \sqrt{\frac{8}{\pi}} \left(\frac{1}{4} + \sum_{i=1}^N \left(1 - \frac{1}{2} D_1(\mathcal{L}(X_i)) \right) \right)^{-1/2}.$$

The following two theorems are our main contributions in this section. To illustrate their use, we apply them in a simple setting in Proposition 3.8 at the end of this section.

Theorem 3.6. *Let (X, X') be an exchangeable pair and let $W := W(X)$ and $W' := W(X')$ take values on the integers. Define*

$$Q_m(x) = \mathbb{P}[W' = W + m | X = x]$$

and $q_m = \mathbb{E}Q_m(X) = \mathbb{P}[W' = W + m]$. Then, for every positive integer m ,

$$D_{1,m}(\mathcal{L}(W)) \leq \frac{\sqrt{\text{Var } Q_m(X)} + \sqrt{\text{Var } Q_{-m}(X)}}{q_m}.$$

Proof. To prove the first assertion, we must bound $|\mathbb{E}\Delta_m g(W)|$ for all g with norm no greater than one. To this end, exchangeability implies that for all bounded functions g

$$\begin{aligned} 0 &= q_m^{-1} \mathbb{E}\{ \mathbb{I}[W' = W + m]g(W') - \mathbb{I}[W' = W - m]g(W) \} \\ &= q_m^{-1} \mathbb{E}\{ Q_m(X)g(W + m) - Q_{-m}(X)g(W) \}, \end{aligned} \tag{3.4}$$

so that

$$\begin{aligned} &|\mathbb{E}\Delta_m g(W)| \\ &= |\mathbb{E}\{ (1 - q_m^{-1} Q_m(X))g(W + m) - (1 - q_m^{-1} Q_{-m}(X))g(W) \}| \\ &\leq \frac{\sqrt{\mathbb{E}\{g(W + m)^2\} \text{Var } Q_m(X)} + \sqrt{\mathbb{E}\{g(W)^2\} \text{Var } Q_{-m}(X)}}{q_m}, \end{aligned} \tag{3.5}$$

where in the inequality we use first the triangle inequality and then Cauchy–Schwarz. Taking the supremum over g with $\|g\|_\infty \leq 1$ in (3.5) proves the theorem. \square

Theorem 3.6 is inspired by Stein’s method of exchangeable pairs as used by Chatterjee, Diaconis and Meckes [9] and Röllin [30]. Our next result extends and embellishes Theorem 3.6.

Theorem 3.7. *Let (X, X', X'') be three consecutive steps of a reversible Markov chain in equilibrium. Let W and W' be as in Theorem 3.6 and, in addition, $W'' := W(X'')$. Define*

$$Q_{m_1, m_2}(x) = \mathbb{P}[W' = W + m_1, W'' = W' + m_2 | X = x].$$

Then, for every positive integer m ,

$$\begin{aligned} D_{2,m}(\mathcal{L}(W)) &\leq \frac{1}{q_m^2} (2 \operatorname{Var} Q_m(X) + \mathbb{E}|Q_{m,m}(X) - Q_m(X)|^2) \\ &\quad + 2 \operatorname{Var} Q_{-m}(X) + \mathbb{E}|Q_{-m,-m}(X) - Q_{-m}(X)|^2. \end{aligned}$$

Proof. Similar to the proof of Theorem 3.6, we want to bound $\mathbb{E}\Delta_m^2 g(W)$ for all g with norm no greater than one. We begin with the trivial equality

$$\mathbb{E}\{\mathbb{I}[W' = W + m, W'' = W' + m]g(W + m)\} \quad (3.6)$$

$$= \mathbb{E}\{\mathbb{I}[W' = W + m, W'' = W' + m]g(W')\}. \quad (3.7)$$

Conditioning on X in (3.6) and on X' in (3.7), the Markov property and reversibility imply

$$\mathbb{E}\{Q_{m,m}(X)g(W + m)\} = \mathbb{E}\{Q_m(X)Q_{-m}(X)g(W)\},$$

and similarly

$$\mathbb{E}\{Q_{-m,-m}(X)g(W)\} = \mathbb{E}\{Q_m(X)Q_{-m}(X)g(W + m)\}.$$

Using these two equalities coupled with (3.4) we find that for bounded g

$$\begin{aligned} 0 &= \mathbb{E}\{g(W + 2m)(q_m^{-2}Q_{m,m}(X) - 2q_m^{-1}Q_m(X))\} \\ &\quad - 2\mathbb{E}\{g(W + m)(q_m^{-2}Q_m(X)Q_{-m}(X) - q_m^{-1}Q_m(X) - q_m^{-1}Q_{-m}(X))\} \\ &\quad + \mathbb{E}\{g(W)(q_m^{-2}Q_{-m,-m}(X) - 2q_m^{-1}Q_{-m}(X))\}. \end{aligned}$$

It is now not hard to see that

$$\begin{aligned} \mathbb{E}\Delta_m^2 g(W) &= \mathbb{E}g(W + 2m) - 2\mathbb{E}g(W + m) + \mathbb{E}G(W) \\ &= \mathbb{E}\{g(W + 2m)((1 - q_m^{-1}Q_m(X))^2 + q_m^{-2}(Q_{m,m}(X) - Q_m(X)^2))\} \\ &\quad - 2\mathbb{E}\{g(W + m)(1 - q_m^{-1}Q_m(X))(1 - q_m^{-1}Q_{-m}(X))\} \\ &\quad + \mathbb{E}\{g(W)((1 - q_m^{-1}Q_{-m}(X))^2 + q_m^{-2}(Q_{-m,-m}(X) - Q_{-m}(X)^2))\}. \end{aligned}$$

The theorem now follows by taking the supremum over g with norm no greater than one and applying the triangle inequality and Cauchy–Schwarz. \square

To better understand how Theorems 3.6 and 3.7 work in practice, we derive the following result.

Proposition 3.8. *If $W \sim \text{Bi}(n, p)$, then*

$$D_2(\mathcal{L}(W)) \leq \frac{1}{n} \left(\frac{2p+1}{1-p} + \frac{2(1-p)+1}{p} \right).$$

Proof. Retaining the notation above, we define the following Markov chain on sequences of zeros and ones of length n , reversible with respect to the Bernoulli product measure. At each step in the chain, a coordinate is selected uniformly at random and resampled. Let X, X', X'' be three consecutive steps in this chain in stationary and $W (= W(X)), W', W''$ be the number of ones in these 0 – 1 configurations. We find

$$Q_1(X) = \frac{n-W}{n}p \quad \text{and} \quad Q_{-1}(X) = \frac{W}{n}(1-p),$$

since, for example, in order for the number of ones to increase by one from X , a zero must be selected (with probability $(n - W)/n$) and must be resampled as a one (with probability p). Similarly, we have

$$Q_{1,1}(X) = \frac{(n-W)(n-W-1)}{n^2}p^2,$$

$$Q_{-1,-1}(X) = \frac{W(W-1)}{n^2}(1-p)^2,$$

since in order for the number of ones to increase by one from X and then again from X' , at both steps a zero must be selected (with probability $((n - W)/n)((n - W - 1)/n$) and then at both steps the selected coordinate must be resampled as a one (with probability p^2). Now, basic properties of the binomial distribution show

$$q_1 = \mathbb{E}Q_1(X) = p(1-p),$$

$$\text{Var}(Q_1(X)) = \frac{p^3(1-p)}{n}, \quad \text{Var}(Q_{-1}(X)) = \frac{(1-p)^3p}{n},$$

$$\mathbb{E}|Q_{1,1}(X) - Q_1(X)^2| = \frac{p^2}{n^2}\mathbb{E}(n-W) = \frac{p^2(1-p)}{n},$$

$$\mathbb{E}|Q_{-1,-1}(X) - Q_{-1}(X)^2| = \frac{(1-p)^2}{n^2}\mathbb{E}W = \frac{(1-p)^2p}{n},$$

and the result follows after putting these values into Theorem 3.7 and simplifying. □

4. Applications

Because we are going to work in the total variation and local limit metrics, we need to use a discrete analog of the normal distribution. We use the *translated Poisson* distribution, but any distribution such that an analog of Lemma 4.1 below holds would also work in the examples below (for example, any standard discretization of the normal distribution). We say

that the random variable Z has the translated Poisson distribution, denoted $Z \sim \text{TP}(\mu, \sigma^2)$, if $Z - \lfloor \mu - \sigma^2 \rfloor \sim \text{Po}(\sigma^2 + \gamma)$, where $\gamma = \mu - \sigma^2 - \lfloor \mu - \sigma^2 \rfloor$. Note that $\mathbb{E}Z = \mu$ and $\sigma^2 \leq \text{Var } Z \leq \sigma^2 + 1$. The translated Poisson distribution is a Poisson distribution shifted by an integer to closely match a given mean and variance; see Röllin [30] for basic properties and applications.

The following lemma essentially states that we can use the translated Poisson distribution as a discrete substitute for the normal distribution and also provides bounds on the appropriate smoothing terms.

Lemma 4.1. *If $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, then as $\sigma \rightarrow \infty$,*

$$D_{k,m}(\text{TP}(\mu, \sigma^2)) = O(\sigma^{-k}), \tag{4.1}$$

$$d_K(\text{TP}(\mu, \sigma^2), N(\mu, \sigma^2)) = O(\sigma^{-1}), \tag{4.2}$$

$$d_W(\text{TP}(\mu, \sigma^2), N(\mu, \sigma^2)) = O(1) \tag{4.3}$$

and

$$\sup_{k \in \mathbb{Z}} \left| \text{TP}(\mu, \sigma^2)\{k\} - \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(k - \mu)^2}{2\sigma^2}\right) \right| = O(\sigma^{-2}). \tag{4.4}$$

Remark 4.1. Let us make a few clarifying remarks about Lemma 4.1 and its use in what follows. First note that as the proof below shows, the rates obtained in Lemma 4.1 hold in general for sums $X_1 + \dots + X_n$ of independent identically distributed random variables with integer support and $D_1(X_1) < 2$. Also, in order to appreciate the Wasserstein bound (4.3), the reader should keep in mind that both distributions in the statement are not standardized and that, for any random variables X and Y and any positive constant c ,

$$d_W(\mathcal{L}(cX), \mathcal{L}(cY)) = cd_W(\mathcal{L}(X), \mathcal{L}(Y)). \tag{4.5}$$

Hence, after scaling the variables in (4.3) by σ^{-1} , the rate becomes the more familiar $O(\sigma^{-1})$. Finally, the statement in (4.4) is just the local limit theorem for the translated Poisson distribution. Such a statement is only informative if the right-hand side of (4.4) is $o(\sigma^{-1})$, because the left-hand side is trivially $O(\sigma^{-1})$. In this section, we will prove bounds for $d_{\text{loc}}(\mathcal{L}(W), \text{TP}(\mu, \sigma^2))$, which are better than $O(\sigma^{-1})$ and therefore, by means of (4.4), will lead to a local limit theorem for W along with a bound on the rate of convergence.

Remark 4.2. Lemma 4.1 can serve as a benchmark for the best possible rates of convergence. For sums of i.i.d. random variables under finite third moment conditions, the Kolmogorov and Wasserstein distances between the normalized random variables and the standard normal distribution are both $O(\sigma^{-1})$, which can be improved only under additional assumptions (such as symmetry) of the involved distributions. Furthermore, if the summands are integer valued and smooth enough, then the local metric distance to a discrete analog of the normal distribution has rate $O(\sigma^{-2})$.

As indicated in the [Introduction](#), our method will typically not yield rates of convergence that are comparable to [Lemma 4.1](#) and those for sums of i.i.d. random variables. So in applications where it is expected the rates should be the same as those for sums of i.i.d. random variables (e.g., magnetization in the Curie–Weiss model at high temperature), our results are likely not optimal. However, in particular for the local limit metric, for which only few results with explicit rates of convergence are known, it is not clear whether one can expect the same rates as those for sums of i.i.d. variables, and so we leave the question of optimality open.

Proof of Lemma 4.1. First, note that since $D_{n,m}(\mathcal{L}(X)) = D_{n,m}(\mathcal{L}(X+l))$ for all integers l , it is enough to prove [\(4.1\)](#) with the translated Poisson distribution replaced by $\text{Po}(\sigma^2 + \gamma)$. We can represent this Poisson distribution as the convolution of k independent Poisson distributions all having mean $(\sigma^2 + \gamma)/k$ and so by [Lemmas 3.2](#) and [3.4](#) we find that for $X \sim \text{Po}((\sigma^2 + \gamma)/k)$,

$$D_{k,m}(\text{TP}(\mu, \sigma^2)) \leq m D_1(\mathcal{L}(X))^k. \tag{4.6}$$

We can represent X as the sum of $\lfloor (\sigma^2 + \gamma)/k \rfloor$ (here assume $\sigma^2 > k$) i.i.d. Poisson variables with means $\lambda_{\sigma,k} \geq 1$. [Lemma 3.5](#) now implies that if $Y_{\sigma,k} \sim \text{Po}(\lambda_{\sigma,k})$ and $D_1(\mathcal{L}(Y_{\sigma,k})) < 2 - \varepsilon$ for some $\varepsilon > 0$ and all σ sufficiently large, then

$$D_1(\mathcal{L}(X)) = O(\sigma^{-1}),$$

which with [\(4.6\)](#) yields [\(4.1\)](#). But it is well known (and easily checked) that for $W \sim \text{Po}(\lambda)$ and any bounded function g , $\lambda \mathbb{E}g(W + 1) = \mathbb{E}\{Wg(W)\}$ and so

$$\begin{aligned} D_1(\mathcal{L}(W)) &= \sup_{\|g\|_\infty \leq 1} |\mathbb{E}g(W + 1) - \mathbb{E}g(W)| \\ &= \frac{1}{\lambda} \sup_{\|g\|_\infty \leq 1} |\mathbb{E}\{(W - \lambda)g(W)\}| \leq \frac{1}{\sqrt{\lambda}}, \end{aligned}$$

where the last inequality follows by Cauchy–Schwarz (using Fourier methods, Barbour, Holst and Janson [\[3\]](#), [Proposition A.2.7](#), in fact show that $D_1(\mathcal{L}(W)) \leq 2/\sqrt{2e\lambda}$). Thus, it is indeed true that $D_1(Y_{\sigma,k}) \leq 1 < 2 - \varepsilon$ and [\(4.1\)](#) is proved.

The remaining properties follow by representing $\text{Po}(\sigma^2 + \gamma)$ as a sum of $\lfloor \sigma^2 \rfloor$ i.i.d. Poisson random variables and using well known theory about sums of independent random variables: [\(4.2\)](#) and [\(4.4\)](#) are respectively [Theorem 4](#) on page 111 and [Theorem 6](#) on page 197 of Petrov [\[27\]](#) and [\(4.3\)](#) is [Corollary 4.2](#) on page 68 of Chen, Goldstein and Shao [\[12\]](#). \square

4.1. Magnetization in the Curie–Weiss model

Let $\beta > 0$, $h \in \mathbb{R}$ and for $s \in \{-1, 1\}^n$ define the Gibbs measure

$$\mathbb{P}(s) = Z^{-1} \exp\left\{ \frac{\beta}{n} \sum_{i < j} s_i s_j + h \sum_i s_i \right\}, \tag{4.7}$$

where Z is the appropriate normalizing constant (we use the letter “ s ” instead of the more commonly used “ σ ” in order to avoid confusion with the notation for variance).

This probability model is referred to as the Curie–Weiss model and a quantity of interest is the magnetization $W = \sum_i s_i$ of the system. The book Ellis [15] provides a good introduction to these models. We use our framework to show total variation and local limit theorems (LLTs) with bounds on the rates of convergence; these are stated in Theorem 4.5 below. We start by stating known limit and approximation results for the Kolmogorov metric, which we will need for our approach.

Theorem 4.2 (Ellis, Newman and Rosen [16], Theorem 2.2). *If s has law given by (4.7) with $0 < \beta < 1$, and $h \in \mathbb{R}$ and $W = \sum_i s_i$, then there is a unique solution m_0 of*

$$m = \tanh(\beta m + h)$$

and as $n \rightarrow \infty$,

$$d_K\left(\mathcal{L}\left(\frac{W - nm_0}{n^{1/2}}\right), N\left(0, \frac{1 - m_0^2}{1 - \beta + \beta m_0^2}\right)\right) \rightarrow 0.$$

Theorem 4.3 (Eichelsbacher and Löwe [14], Theorems 3.3 and 3.7). *If s has law given by (4.7) with $0 < \beta < 1$, and $h = 0$ and $W = \sum_i s_i$, then there is a constant C depending only on β such that*

$$d_K(\mathcal{L}(n^{-1/2}W), N(0, (1 - \beta)^{-1})) \leq Cn^{-1/2};$$

the same bound holds for the Wasserstein metric.

Note that Chen, Fang and Shao [11] have obtained moderate deviation results, which are much sharper than the Berry–Esseen type bounds of Theorem 4.3.

The other ingredient of applying our framework here is to use Theorem 3.7 to bound the necessary smoothing terms. For this purpose, let s as above and s' be a step from s in the following reversible Markov chain: at each step of the chain a site from the n possible sites is chosen uniformly at random and then the spin at that site is resampled according to the Gibbs measure (4.7) conditional on the value of the spins at all other sites. Let $W = \sum_{i=1}^n s_i$ and $W' = \sum_{i=1}^n s'_i$ and note that (W, W') is an exchangeable pair. Finally, define

$$Q_m = \mathbb{P}[W' = W + m | s],$$

$q_m = \mathbb{E}Q_m$, and

$$Q_{m_1, m_2} = \mathbb{P}[W' = W + m_1, W'' = W' + m_2 | s],$$

where W'' is obtained from W' in the same way that W' is obtained from W (i.e., (W, W', W'') are the magnetizations in three consecutive steps in the stationary Markov chain described above). We have the following result, proved at the end of this section.

Lemma 4.4. *If $0 < \beta < 1$, $h \in \mathbb{R}$ and $M = \frac{1}{n} \sum_{i=1}^n s_i$, then there is a unique solution m_0 to*

$$m = \tanh(\beta m + h),$$

and for $k = \pm 2$,

$$\left| Q_k - \frac{1 - m_0^2}{4} \right| \leq C \left(|M - m_0| + \frac{1}{n} \right), \tag{4.8}$$

$$|Q_{k,k} - Q_k^2| = O(n^{-1}), \tag{4.9}$$

$$\left| q_k - \frac{1 - m_0^2}{4} \right| = O(n^{-1/2}), \quad \text{Var}(Q_k) = O(n^{-1}) \tag{4.10}$$

and

$$D_{2,2}(W) = O(n^{-1}). \tag{4.11}$$

We can now put these pieces together to obtain total variation and local limit convergence theorems with bounds on the rates for the magnetization.

Theorem 4.5. *Let s have law given by (4.7), $W = \sum_i s_i$, and let $\delta = \delta(n) = (1 - (-1)^n)/2$. For $0 < \beta < 1$ and $h = 0$, there is a constant C that depends only on β such that*

$$d_{\text{loc}} \left(\mathcal{L}((W + \delta)/2), \text{TP} \left(0, \frac{n}{4(1 - \beta)} \right) \right) \leq C n^{-3/4},$$

$$d_{\text{TV}} \left(\mathcal{L}((W + \delta)/2), \text{TP} \left(0, \frac{n}{4(1 - \beta)} \right) \right) \leq C n^{-1/3}.$$

If $0 < \beta < 1$, $h \in \mathbb{R}$, and m_0 is as in Theorem 4.2, then

$$d_{\text{loc}} \left(\mathcal{L} \left(\frac{W + \delta}{2} \right), \text{TP} \left(\frac{m_0}{2}, \frac{n(1 - m_0^2)}{4(1 - \beta + \beta m_0^2)} \right) \right) = o(n^{-1/2})$$

as $n \rightarrow \infty$.

Proof. The theorem follows from (ii), (iii) and (iv) of Theorem 2.2 with $m = 1$ and $l = 2$, Lemma 4.1, Theorems 4.3 and 4.2, and the bounds on the smoothing terms in Lemma 4.4. \square

Remark 4.3. In the critical case where $\beta = 1$, optimal bounds on the Kolmogorov and Wasserstein distances between the magnetization (appropriately normalized) and its non-normal limiting distribution have been obtained by Eichelsbacher and Löwe [14], Theorem 3.3, and Chatterjee and Shao [10], Theorem 2.1. In fact, the smoothing bounds of Lemma 4.4 can be shown to apply to this case with $h = 0$ and an appropriate analog of Lemma 4.1 also holds for a discretization of the non-normal limiting distribution. And these two facts can be used to prove new bounds on the total variation distance between the magnetization and a discrete version of this limiting

distribution (although after working out the details, we are not able to obtain meaningful local metric results). However, we omit this result due to the inappropriate amount of space it would take for a precise formulation.

Proof of Lemma 4.4. We only consider $k = 2$, the case $k = -2$ being similar. An easy calculation shows that

$$\mathbb{P}(s'_i = 1 | (s_j)_{j \neq i}) = \frac{\exp\{(\beta/n) \sum_{j \neq i} s_j + h\}}{\exp\{(\beta/n) \sum_{j \neq i} s_j + h\} + \exp\{-(\beta/n) \sum_{j \neq i} s_j - h\}}.$$

Denoting $m_i := n^{-1} \sum_{j \neq i} s_j$, we have

$$Q_2 = \frac{1}{n} \sum_{i=1}^n \frac{1-s_i}{2} \mathbb{P}(s'_i = 1 | (s_j)_{j \neq i}) = \frac{1}{n} \sum_{i=1}^n \frac{1-s_i}{2} \frac{\tanh(\beta m_i + h) + 1}{2},$$

since in order for the Markov chain to increase by two, a site in state “−1” must be selected and then changed to “+1”. Now, some simplification shows

$$\begin{aligned} Q_2 &= \frac{1}{4} - \frac{M}{4} + \frac{\tanh(\beta M + h)}{4} (1 - M) \\ &\quad + \frac{1}{4n} \sum_{i=1}^n (1 - s_i) (\tanh(\beta m_i + h) - \tanh(\beta M + h)). \end{aligned}$$

Thus, we find

$$\begin{aligned} \left| Q_2 - \frac{1 - m_0^2}{4} \right| &\leq \frac{1}{4} |M - m_0| + \frac{1}{4} |\tanh(\beta m_0 + h) - \tanh(\beta M + h)| \\ &\quad + \frac{1}{4} |M \tanh(\beta M + h) - m_0 \tanh(\beta m_0 + h)| \\ &\quad + \frac{1}{4n} \sum_{i=1}^n (1 - s_i) |\tanh(\beta m_i + h) - \tanh(\beta M + h)|. \end{aligned}$$

Since $\tanh(x) \in (-1, 1)$ is 1-Lipschitz and $-1 \leq M \leq 1$ the first part of the claim now easily follows.

For the second assertion, note that

$$\begin{aligned} Q_{2,2} &= \frac{1}{16n^2} \sum_{i \neq j} (1 - s_i) (\tanh(\beta m_i + h) + 1) \\ &\quad \times (1 - s_j) (\tanh(\beta m_{i,j} + \beta n^{-1} + h) + 1), \end{aligned}$$

where $m_{i,j} = \beta n^{-1} \sum_{k \neq i,j} s_k$, and also that

$$Q_2^2 = \frac{1}{16n^2} \sum_{i,j} (1 - s_i)(\tanh(\beta m_i + h) + 1)(1 - s_j)(\tanh(\beta m_j + h) + 1).$$

We can now find

$$\begin{aligned} |Q_{2,2} - Q_2^2| &\leq \frac{1}{8n^2} \sum_i (1 - s_i)(\tanh(\beta m_i + h) + 1)^2 \\ &\quad + \frac{1}{8n^2} \sum_{i \neq j} (1 - s_i)(1 - s_j) |\tanh(\beta m_i + h) + 1| \\ &\quad \times |\tanh(\beta m_{i,j} + \beta n^{-1} + h) - \tanh(\beta m_j + h)|. \end{aligned}$$

Straightforward estimates now yield (4.9).

The assertions of (4.10) follow from (4.8) and the fact $\mathbb{E}|M - m_0|^j = O(n^{-j/2})$ which is obtained from standard concentration results; see, for example, Chatterjee [8], Proposition 1.3. Finally, (4.11) follows from (4.8), (4.9), and (4.10) applied to Theorem 3.7. \square

4.2. Isolated vertices in the Erdős–Rényi random graph

In this and the next section, we will derive LLTs for the number of isolated vertices and triangles in the Erdős–Rényi random graph. There do not appear to be many results showing LLTs for random graph variables (and even fewer having error bounds) although one area that has seen activity is showing LLTs for the size of the maximal component in graphs and hypergraphs; see Stepanov [32], Karoński and Łuczak [20] and Behrisch, Coja-Oghlan and Kang [6]. An alternative approach to proving an LLT for the number of isolated vertices in an Erdős–Rényi graph which we do not believe has been pursued would be to use the results of Bender, Canfield and McKay [7] which have detailed formulas for the number of graphs with a given number of vertices and edges and no isolated vertices. To the best of our knowledge, the following results on the number of isolated vertices and number of triangles are new.

Before proceeding, we make a remark to prepare the dedicated reader for the proofs below. Many proofs of limit theorems for random graph variables involve tedious moment calculations. For example, the limit results we use below in our framework: Ruciński [31] uses the method of moments to derive conditions where the number of copies of a “small” subgraph in an Erdős–Rényi graph will be approximately normally distributed, and Barbour, Karoński and Ruciński [4] uses a variation of Stein’s method which in turn relies on moment estimates to show limit theorems for the number of copies of certain subgraphs in an Erdős–Rényi graph; see also the references in these documents. Since our theory relies on bounding means and variances of conditional probabilities, our work below continues this tradition.

We use our framework to show total variation and local limit theorems with bounds on the rates for the number of isolated vertices; this is Theorem 4.8 below. We start by stating known limit and approximation results. Define $G = G(n, p)$ to be a random graph with n vertices where each edge appears with probability p , independent of all other edges.

Theorem 4.6 (Barbour, Karoński and Ruciński [4], Kordecki [21]). *Let $W = W(n, p)$ be the number of isolated vertices of $G(n, p)$, and let \tilde{W} be W normalized to have zero mean and unit variance. Then \tilde{W} converges in distribution to the standard normal if and only if*

$$\lim_{n \rightarrow \infty} n^2 p = \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} (\log(n) - np) = \infty. \quad (4.12)$$

In that case, with $\sigma_n^2 = \text{Var } W$,

$$d_K(\mathcal{L}(\tilde{W}), \Phi) = O(\sigma_n^{-1}).$$

The conditions of convergence was proved by Barbour, Karoński and Ruciński [4], whereas the bounds for the Kolmogorov metric was obtained by Kordecki [21].

The other ingredient of applying our framework here is to use Theorem 3.7 to bound the necessary smoothing terms. We have the following result, proved at the end of this section.

Lemma 4.7. *Let $W = W(n, p)$ be the number of isolated vertices in an Erdős–Rényi graph $G(n, p)$ and $\sigma_n^2 = \text{Var } W$.*

- (i) *If $\lim_{n \rightarrow \infty} (\log(n) - np) = \infty$, and either $\lim_{n \rightarrow \infty} np = \infty$ or $\lim_{n \rightarrow \infty} np = c > 0$, then $\sigma_n^2 \asymp ne^{-np}$ and*

$$D_1(W) = O(\sigma_n^{-1}), \quad D_2(W) = O(\sigma_n^{-2}).$$

- (ii) *If $\lim_{n \rightarrow \infty} np = 0$ and $\lim_{n \rightarrow \infty} n^2 p = \infty$, then $\sigma_n^2 \asymp n^2 p$ and*

$$\begin{aligned} D_1(W) &= O((\sqrt{np} \sigma_n)^{-1}), & D_2(W) &= O((np)^{-1} \sigma_n^{-2}), \\ D_{1,2}(W) &= O(\sigma_n^{-1}), & D_{2,2}(W) &= O(\sigma_n^{-2}). \end{aligned}$$

We now summarize the results of our framework combined with Theorem 4.6 and Lemma 4.7. For two distribution functions F and G with integer support, let

$$d_{\text{loc}}^m(F, G) = \|\Delta \bar{F}^m - \Delta \bar{G}^m\|_{\infty}.$$

Note that $d_{\text{loc}} = d_{\text{loc}}^1$ and recall also the equality given by Lemma 2.3.

Theorem 4.8. *Let $W = W(n, p)$ be the number of isolated vertices in an Erdős–Rényi random graph $G(n, p)$ and \tilde{W} be W normalized to have zero mean and unit variance. With $\mu_n = \mathbb{E}W$ and $\sigma_n^2 = \text{Var}(W)$, we have the following.*

- (i) *If $\lim_{n \rightarrow \infty} (\log(n) - np) = \infty$, and either $\lim_{n \rightarrow \infty} np = \infty$ or $\lim_{n \rightarrow \infty} np = c > 0$,*

$$d_{\text{loc}}(\mathcal{L}(W), \text{TP}(\mu_n, \sigma_n^2)) = O(\sigma_n^{-3/2}).$$

- (ii) *If $\lim_{n \rightarrow \infty} np = 0$ and $\lim_{n \rightarrow \infty} n^2 p = \infty$, then*

$$\begin{aligned} d_{\text{loc}}(\mathcal{L}(W), \text{TP}(\mu_n, \sigma_n^2)) &= O(\sigma_n^{-1} (np^{3/4})^{-1}), \\ d_{\text{loc}}^2(\mathcal{L}(W), \text{TP}(\mu_n, \sigma_n^2)) &= O(\sigma_n^{-3/2}). \end{aligned}$$

Proof. The result follows from (ii) of Theorem 2.2 with $l = 2$, using the known rates stated above in Theorem 4.6 coupled with Lemma 4.1, and bounds on the smoothing quantities provided by Lemma 4.7. \square

Remark 4.4. The bounds on the smoothing quantities presented below can be written in terms of n and p , so that the asymptotic results of the theorem can be written explicitly whenever the bounds on $d_K(\tilde{W}, \Phi)$ are also explicit.

Remark 4.5. The second case of the theorem is interesting and deserves elaboration. In some regimes, our bounds do not imply a LLT in the natural lattice of span one (e.g., $p \asymp n^{-\alpha}$, where $\alpha > 4/3$), but we can obtain a useful bound on the rate of convergence in the d_{loc}^2 distance. This implies that the approximation is better by averaging the probability mass function of W over two neighboring integers and then comparing this value to its analog for the normal density. One explanation for this phenomenon is that in such a regime, the graph $G(n, p)$ will be extremely sparse so that parity of W will be dominated by the number of isolated edges. In other words, with some significant probability, $n - W$ will be approximately equal to twice the number of isolated edges, in which case we would not expect the normal density to be a good approximation for each point on the integer lattice.

Proof of Lemma 4.7. The stated order of the variance follows from

$$\sigma_n^2 = n(1 - p)^{n-1} [1 + (np - 1)(1 - p)^{n-2}],$$

which follows easily after representing W as a sum of indicators; see also the moment information given below.

To prove the bounds on the smoothing terms, we apply Theorems 3.6 and 3.7. For this purpose, let $G(n, p)$ as above and $G'(n, p)$ be a step from $G(n, p)$ in the following reversible Markov chain: from a given graph G the chain moves to G' by choosing two vertices uniformly at random and resampling the “edge” between them. Let $W = W(n, p)$ be the number of isolated vertices of the Erdős–Rényi graph $G = G(n, p)$ and W' be the number of isolated vertices after one step in the chain from G , so (W, W') is an exchangeable pair. Finally, define

$$Q_m = \mathbb{P}[W' = W + m | \sigma],$$

$q_m = \mathbb{E}Q_m$, and

$$Q_{m_1, m_2} = \mathbb{P}[W' = W + m_1, W'' = W' + m_2 | \sigma],$$

where W'' is obtained from W' in the same way that W' is obtained from W (i.e. (W, W', W'') are the magnetizations in three consecutive steps in the stationary Markov chain described above).

In order to compute the terms needed to apply Theorems 3.6 and 3.7, we need some auxiliary random variables. Let W_k be the number of vertices of degree k in G (so $W_0 \equiv W$), and E_2 be the number of connected pairs of vertices each having degree one (i.e., E_2 is the number of isolated

edges). We have

$$\begin{aligned}
 Q_1(G) &= \frac{(W_1 - 2E_2)}{\binom{n}{2}}(1 - p), & Q_{-1}(G) &= \frac{W(n - W)}{\binom{n}{2}}p, \\
 Q_{1,1}(G) &= \frac{2\binom{W_1 - 2E_2}{2}}{\binom{n}{2}^2}(1 - p)^2, & Q_{-1,-1}(G) &= \frac{4\binom{W}{2}\binom{n - W + 1}{2}}{\binom{n}{2}^2}p^2, \\
 Q_2(G) &= \frac{E_2}{\binom{n}{2}}(1 - p), & Q_{-2}(G) &= \frac{\binom{W}{2}}{\binom{n}{2}}p, \\
 Q_{2,2}(G) &= \frac{2\binom{E_2}{2}}{\binom{n}{2}^2}(1 - p)^2, & Q_{-2,-2}(G) &= \frac{\binom{W}{2}\binom{W - 2}{2}}{\binom{n}{2}^2}p^2.
 \end{aligned}$$

These equalities are obtained through straightforward considerations. For example, in order for one step in the chain to increase the number of isolated vertices by one, an edge of G must be chosen that has exactly one end vertex of degree one, and then must be removed upon resampling. For one step in the chain to decrease the number of isolated vertices by one, an isolated vertex must be connected to a vertex with positive degree.

From this point, the lemma will follow after computing the pertinent moment information needed to apply Theorems 3.6 and 3.7. By considering appropriate indicator functions, it is an elementary combinatorial exercise to obtain

$$\begin{aligned}
 \mathbb{E}W_1 &= 2\binom{n}{2}p(1 - p)^{n-2}, & \mathbb{E}E_2 &= \binom{n}{2}p(1 - p)^{2n-4}, \\
 \mathbb{E}W_1^2 &= 2\binom{n}{2}p(1 - p)^{n-2} + 2p\binom{n}{2}[(1 - p)^{2n-4} + p(n - 2)^2(1 - p)^{2n-5}], \\
 \mathbb{E}E_2^2 &= \binom{n}{2}p(1 - p)^{2n-4} + 6\binom{n}{4}p^2(1 - p)^{4n-12}, \\
 \mathbb{E}W_1E_2 &= \binom{n}{2}p(1 - p)^{2n-4}[(n - 2)(n - 3)p(1 - p)^{n-4} + 2],
 \end{aligned}$$

which will yield the results for negative jumps, and

$$\begin{aligned}
 \mathbb{E}W &= n(1 - p)^{n-1}, & \mathbb{E}W^2 &= n(1 - p)^{n-1} + 2\binom{n}{2}(1 - p)^{2n-3}, \\
 \mathbb{E}W^3 &= n(1 - p)^{n-1} + 6\binom{n}{2}(1 - p)^{2n-3} + 6\binom{n}{3}(1 - p)^{3n-6}, \\
 \mathbb{E}W^4 &= n(1 - p)^{n-1} + 14\binom{n}{2}(1 - p)^{2n-3} \\
 &\quad + 36\binom{n}{3}(1 - p)^{3n-6} + 24\binom{n}{4}(1 - p)^{4n-10},
 \end{aligned}$$

which will yield the results for the positive jumps. Theorems 3.6 and 3.7 now give the desired rates. As an example of these calculations, note that

$$\frac{\text{Var } Q_1(G)}{q_1^2} = \frac{\mathbb{E}W_1^2 - 4\mathbb{E}W_1E_2 + 4\mathbb{E}E_2^2 - (\mathbb{E}W_1 - 2\mathbb{E}E_2)^2}{(\mathbb{E}W_1 - 2\mathbb{E}E_2)^2}, \tag{4.13}$$

which after the dust settles is $O(n^{-1}e^{np})$ in case (i) of the theorem. Similarly, since $q_1 = \mathbb{E}Q_{-1}$, we have

$$\frac{\text{Var } Q_{-1}(G)}{q_1^2} = \frac{\mathbb{E}W^4 - 2n\mathbb{E}W^3 + n^2\mathbb{E}W^2 - (n\mathbb{E}W - \mathbb{E}W^2)^2}{(n\mathbb{E}W - \mathbb{E}W^2)^2}, \tag{4.14}$$

which is again $O(n^{-1}e^{np})$ in case (i). Theorem 3.6 implies that $D_1(W)$ is bounded above by the sum of the square roots of the terms in (4.13) and (4.14) so that in case (i),

$$D_1(W) = O(n^{-1/2}e^{(np)/2}) = O(\sigma_n^{-1}).$$

For the second part of (i), note that

$$\frac{\mathbb{E}|Q_{1,1} - Q_1^2|}{q_1^2} \leq \frac{\mathbb{E}W_1 + 2\mathbb{E}E_2}{(\mathbb{E}W_1 - 2\mathbb{E}E_2)^2},$$

which is $O(n^{-2}p^{-1}e^{np})$ in case (i) and

$$\frac{\mathbb{E}|Q_{-1,-1} - Q_{-1}^2|}{q_1^2} = \frac{\mathbb{E}(W(n - W)|n - 2W + 1|)}{(n\mathbb{E}W - \mathbb{E}W^2)^2} \leq \frac{n + 1}{n\mathbb{E}W - \mathbb{E}W^2},$$

which is $O(n^{-1}e^{np})$ in case (i), so that we have

$$D_2(W) = O(n^{-1}e^{np}) = O(\sigma_n^{-2}).$$

This proves (i); the remaining bounds are similar and omitted for the sake of brevity. □

4.3. Triangles in the Erdős–Rényi random graph

In this section, we use our framework to first obtain a new bound on the rate of convergence in the total variation distance between the normal distribution and the number of triangles in an Erdős–Rényi random graph. We then use this new rate to obtain a local limit theorem for this example. As in Section 4.2, define $G = G(n, p)$ to be a random graph with n vertices where each edge appears with probability p , independent of all other edges. From this point, we have the following theorem.

Theorem 4.9 (Ruciński [31], Barbour, Karoński and Ruciński [4]). *Let $W = W(n, p)$ be the number of triangles of $G(n, p)$, and let \tilde{W} be W normalized to have zero mean and unit variance.*

Then \tilde{W} converges to the standard normal if and only if

$$\lim_{n \rightarrow \infty} np = \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} n^2(1 - p) = \infty.$$

In that case, with $\sigma_n^2 = \text{Var } W$,

$$d_W(\mathcal{L}(\tilde{W}), \Phi) = O(\sigma_n^{-1}).$$

The other ingredient of applying our framework here is to use Theorem 3.7 to bound the necessary smoothing terms. We have the following result, proved at the end of this section.

Lemma 4.10. *Let $W = W(n, p)$ be the number of triangles in an Erdős–Rényi random graph $G(n, p)$. If $n^\alpha p \rightarrow c > 0$ with $1/2 \leq \alpha < 1$ then $\text{Var}(W) \asymp n^3 p^3$ and*

$$D_1(W) = O(\sigma_n^{-1}), \quad D_2(W) = O(\sigma_n^{-2}). \tag{4.15}$$

We now summarize the results derived from the bound of Theorem 4.9 coupled with our theory above.

Theorem 4.11. *Let $W = W(n, p)$ be the number of triangles in an Erdős–Rényi random graph $G(n, p)$. If $n^\alpha p \rightarrow c > 0$ with $1/2 \leq \alpha < 1$ then with $\mu_n = \mathbb{E}W$ and $\sigma_n^2 := \text{Var}(W)$, we have*

$$d_{\text{TV}}(\mathcal{L}(W), \text{TP}(\mu_n, \sigma_n^2)) = O(n^{-(1-\alpha)})$$

and

$$d_{\text{loc}}(\mathcal{L}(W), \text{TP}(\mu_n, \sigma_n^2)) = O(\sigma_n^{-1} n^{-(1-\alpha)/2}).$$

Proof. The result follows from (iv) and then (i) (or (iii)) of Theorem 2.2 with $l = 2$ and $m = 1$, using the known rates stated above in Theorem 4.9 coupled with Lemma 4.1 and bounds on the smoothing quantities provided by Lemma 4.10. □

Remark 4.6. It is worthwhile noting that we obtain the LLT only for those values of α for which we have $\mathbb{E}W \asymp \text{Var } W$. In contrast, if $0 < \alpha < 1/2$, we have that $\mathbb{E}W \asymp n^{3-3\alpha}$, whereas $\text{Var } W \asymp n^{4-5\alpha} \gg \mathbb{E}W$. It is not clear if this is an artifact of our method or if a standard LLT does not hold in this regime; cf. Remark 4.5 following Theorem 4.8.

In order to prove Lemma 4.10, we will apply Theorems 3.6 and 3.7 by constructing a Markov chain on graphs with n vertices which is reversible with respect to the law of $G(n, p)$. From a given graph G , define a step in the chain to G' by choosing two vertices of G uniformly at random and independently resampling the “edge” between them. It is clear that this Markov chain is reversible with respect to the distribution of $G(n, p)$. We are now in a position to compute the terms needed to apply Theorems 3.6 and 3.7.

Lemma 4.12. *Let (W, W') be the number of triangles in the exchangeable pair of Erdős–Rényi graphs (G, G') as defined above. If $Q_1(G) = \mathbb{P}[W' = W + 1|G]$, then*

$$\begin{aligned} \text{Var } Q_1(G) &\leq \frac{\binom{n-2}{n}}{\binom{n}{2}} p^4(1-p)(1-p^2)^{n-3}(1-p^2(1-p)(1-p^2)^{n-3}) \end{aligned} \tag{4.16}$$

$$+ \frac{4\binom{n-2}{2}}{\binom{n}{2}} p^5(1-p)^2((1-2p^2+p^3)^{n-4} - p(1-p^2)^{2n-6}) \tag{4.17}$$

$$+ \frac{4\binom{n-2}{2}}{\binom{n}{2}} p^5(1-p)^2(1-p^2)^{2n-8}(1-p-p(1-p^2)^2) \tag{4.18}$$

$$+ \frac{12\binom{n-2}{3}}{\binom{n}{2}} p^6(1-p)^2((1-p)^{n-3}(1+p-p^2)^{n-5} - (1-p^2)^{2n-6}) \tag{4.19}$$

$$+ \frac{12\binom{n-2}{3}}{\binom{n}{2}} p^6(1-p)^2(1-p^2)^{2n-9}(-2p+4p^2-3p^4+p^6) \tag{4.20}$$

$$+ \frac{3\binom{n-2}{3}}{\binom{n}{2}} p^6(1-p)^2(1-p^2)^{2n-10}(4p^3-7p^4+4p^6-p^8) \tag{4.21}$$

$$+ \frac{12\binom{n-2}{4}}{\binom{n}{2}} p^6(1-p)^2(1-p^2)^{2n-10}(4p^3-7p^4+4p^6-p^8). \tag{4.22}$$

Proof. Let $X_{i,j}$ be the indicator that there is an edge between vertices i and j and $V_i^{k,j} := X_{i,j}X_{i,k}$ be the indicator that there is a V-star on the vertices $\{i, j, k\}$ with elbow i . We easily find

$$Q_1(G) = \frac{p}{\binom{n}{2}} \sum_{\{j,k\}} \sum_{i \neq j,k} Y_i^{j,k},$$

where we define the indicator variables

$$Y_i^{j,k} = (1 - X_{j,k})V_i^{j,k} \prod_{l \neq i,j,k} (1 - V_l^{j,k}).$$

From this point, we note that the variance of $Q_1(G)$ is a sum of covariance terms times $p^2/\binom{n}{2}^2$. For fixed i, j , and k , there are $3\binom{n}{3}$ terms of the form $\text{Cov}(Y_i^{j,k}, Y_u^{s,t})$, where we are including $\text{Var}(Y_i^{j,k})$. In order to compute this sum, we will group these covariance terms with respect to the number of indices $Y_i^{j,k}$ and $Y_u^{s,t}$ share, which will yield the lemma after computing their covariances.

As an example of the type of calculation involved in computing these covariance terms, note that

$$\mathbb{E}Y_i^{j,k} = p^2(1-p)(1-p^2)^{2n-6},$$

so that

$$\text{Var } Y_i^{j,k} = p^2(1-p)(1-p^2)^{2n-6}(1-p^2(1-p)(1-p^2)^{2n-6}).$$

Furthermore, for $j \neq s$, we find that

$$\mathbb{E}\{Y_i^{j,k}Y_i^{s,k}\} = p^3(1-p)^2((1-p)^3 + 3p(1-p)^2 + p^2(1-p))^{n-4}.$$

Below we focus on carefully spelling out the number and types of covariance terms that contribute to the variance of $Q_1(G)$ and leave to the reader detailed calculations similar to those above.

If $\{i, j, k\} = \{u, s, t\}$ and $Y_i^{j,k} \neq Y_u^{s,t}$, then $\mathbb{E}\{Y_i^{j,k}Y_u^{s,t}\} = 0$, so the corresponding covariance term is negative, which we bound above by zero. In the case that $Y_i^{j,k} = Y_u^{s,t}$, we obtain a variance term which corresponds to (4.16) in our bound.

Assume now that $\{i, j, k\}$ and $\{u, s, t\}$ have exactly two elements in common and consider which indices are equal. In the cases that $Y_u^{s,t}$ is equal to $Y_u^{j,k}$, $Y_j^{k,t}$, $Y_u^{i,k}$, $Y_u^{i,j}$ or $Y_k^{j,t}$, then $\mathbb{E}\{Y_i^{j,k}Y_u^{s,t}\} = 0$, so the corresponding covariance term is negative, which we bound above by zero. The two remaining cases to consider are $Y_u^{s,t} = Y_i^{s,k}$ which contribute $2(n-3)$ equal covariance terms leading to (4.17), and $Y_u^{s,t} = Y_j^{i,t}$ which also contribute $2(n-3)$ equal covariance terms leading to (4.18).

Assume $\{i, j, k\}$ and $\{u, s, t\}$ have exactly one element in common; we have four cases to consider. There are $2(n-3)(n-4)$ covariance terms of the basic form $Y_u^{s,t} = Y_u^{j,t}$ which leads to (4.19), there are $2(n-3)(n-4)$ covariance terms of the basic forms $Y_u^{s,t} = Y_u^{i,t}$ or $Y_u^{s,t} = Y_j^{s,t}$ which leads to (4.20), and there are $\binom{n-3}{2}$ terms of the form $Y_u^{s,t} = Y_i^{s,t}$ which yields (4.21).

Finally, if $\{i, j, k\}$ and $\{u, s, t\}$ are distinct sets, of which we have $3\binom{n-3}{3}$ ways of obtaining $Y_u^{s,t}$, the corresponding covariance terms contribute (4.22) to the bound. \square

Lemma 4.13. *Let (W, W') be the number of triangles in the exchangeable pair of Erdős–Rényi graphs (G, G') as defined above. If $Q_{-1}(G) = \mathbb{P}[W' = W - 1|G]$, then*

$$\begin{aligned} \text{Var } Q_{-1}(G) &\leq \frac{\binom{n-2}{2}}{\binom{n}{2}} p^3(1-p)^2(1-p^2)^{n-3}(1-p^3(1-p^2)^{n-3}) \\ &\quad + \frac{2\binom{n-2}{2}}{\binom{n}{2}} p^3(1-p)^2((1-2p^2+p^3)^{n-3} - p^3(1-p^2)^{2n-6}) \\ &\quad + \frac{4\binom{n-2}{2}}{\binom{n}{2}} p^5(1-p)^2((1-p)(1-2p+p^3)^{n-4} - p(1-p^2)^{2n-6}) \end{aligned}$$

$$\begin{aligned}
 &+ \frac{4\binom{n-2}{2}}{\binom{n}{2}} p^5 (1-p)^2 (1-p^2)^{2n-8} (1-p-p(1-p^2)^2) \\
 &+ \frac{12\binom{n-2}{3}}{\binom{n}{2}} p^6 (1-p)^2 ((1-p)^3 (1-2p+p^3)^{n-5} - (1-p^2)^{2n-6}) \\
 &+ \frac{12\binom{n-2}{3}}{\binom{n}{2}} p^6 (1-p)^2 (1-p^2)^{2n-9} ((1-p)^2 - (1-p^2)^3) \\
 &+ \frac{3\binom{n-2}{3}}{\binom{n}{2}} p^6 (1-p)^2 (1-p^2)^{2n-10} (4p^3 - 7p^4 + 4p^6 - p^8) \\
 &+ \frac{12\binom{n-2}{4}}{\binom{n}{2}} p^6 (1-p)^2 (1-p^2)^{2n-10} (4p^3 - 7p^4 + 4p^6 - p^8).
 \end{aligned}$$

Proof. As in the proof of Lemma 4.12, let $X_{i,j}$ be the indicator that there is an edge between vertices i and j and $V_i^{k,j} := X_{i,j} X_{i,k}$ be the indicator that there is a V -star on the vertices $\{i, j, k\}$ with elbow i . We easily find

$$Q_{-1}(G) = \frac{(1-p)}{\binom{n}{2}} \sum_{\{j,k\}} \sum_{i \neq j,k} X_{j,k} V_i^{j,k} \prod_{l \neq i,j,k} (1 - V_l^{j,k}),$$

and from this point, the proof is very similar to the proof of Lemma 4.12. □

We are now in the position to prove Lemma 4.10.

Proof of Lemma 4.10. Recall the notation of Lemmas 4.12 and 4.13. We have the following easy facts:

- (i) $\mathbb{E}W = \binom{n}{3} p^3$,
- (ii) $\sigma_n^2 := \text{Var } W = \binom{n}{3} (p^3(1-p^3) + 3(n-3)p^5(1-p))$,
- (iii) $q_1 = \mathbb{E}Q_1(G) = (n-2)p^3(1-p)(1-p^2)^{n-3}$;

the second item yields the assertion about the rate of $\text{Var}(W)$. The first bound in (4.15) now follows from Theorem 3.6 after noting that for (W, W') the number of triangles in the exchangeable pair of Erdős–Rényi graphs (G, G') as defined above and $n^\alpha p \rightarrow c > 0$ for $1/2 \leq \alpha < 1$, then

$$\text{Var } Q_1(G) = O(p^5), \quad \text{Var } Q_{-1}(G) = O(p^3/n),$$

which follows easily from Lemmas 4.12 and 4.13 above.

In order to prove the second bound in (4.15), we will apply Theorem 3.7 with $G = G(n, p)$ an Erdős–Rényi random graph, G' obtained by taking a step from G in the Markov chain (reversible with respect to the law of $G(n, p)$) defined previously, and G'' obtained as a step from G' in the same Markov chain. Setting (W, W', W'') to be the number of triangles in the graphs

(G, G', G'') , and defining (as per Theorem 3.7)

$$Q_{i,i}(G) = \mathbb{P}[W'' = W' + i, W' = W + i | G],$$

it is easy to see that

$$Q_{1,1}(G) = \frac{p^2}{\binom{n}{2}} \sum_{\{j,k\}} \sum_{i \neq j,k} Y_i^{j,k} \sum_{\{s,t\} \neq \{j,k\}} \sum_{u \neq s,t} Y_u^{s,t},$$

where as in the proof of Lemma 4.12 we define

$$Y_i^{j,k} = (1 - X_{j,k})X_{i,j}X_{i,k} \prod_{l \neq i,j,k} (1 - X_{l,j}X_{l,k}).$$

From this point, we find

$$\mathbb{E}|Q_{1,1}(G) - Q_1(G)^2| = \frac{p}{\binom{n}{2}} q_1,$$

since for fixed $\{j, k\}$, only one of the set $\{Y_i^{j,k}\}_{i=1}^n$ can be non-zero. A similar analysis shows

$$\mathbb{E}|Q_{-1,-1}(G) - Q_{-1}(G)^2| = \frac{1-p}{\binom{n}{2}} q_1$$

and the second bound in (4.15) now follows from Theorem 3.7 after collecting the pertinent facts above. □

4.4. Embedded sum of independent random variables

We consider the case where W has an embedded sum of independent random variables. This setting has been the most prominent way to prove LLTs by probabilistic arguments; see, for example, Davis and McDonald [13], Röllin [29], Barbour [1], Behrisch, Coja-Oghlan and Kang [6] and Penrose and Peres [25]. In this case, our theory can be used to obtain bounds on the rates for an LLT using previously established bounds on rates of convergence in other metrics.

Let W be an integer valued random variable with variance σ^2 and let \mathcal{F} be some σ -algebra. Assume that W allows for a decomposition of the form

$$W = Y + \sum_{i=1}^N Z_i, \tag{4.23}$$

where N is \mathcal{F} -measurable, and where, conditional on \mathcal{F} , we have that Y, Z_1, \dots, Z_N are all independent of each other. Note that in what follows, the distribution of Y is not relevant.

Theorem 4.14. *Let $W = W(\sigma)$ be a family of integer valued random variables satisfying (4.23) and with $\text{Var } W = \sigma^2$. Assume there are constants u and β , independent of σ^2 , such that, conditional on \mathcal{F} ,*

$$0 < u \leq 1 - \frac{1}{2} D_1(Z_i) \tag{4.24}$$

for all $1 \leq i \leq N$, and such that

$$\mathbb{P}[N < \beta\sigma^2] = O(\sigma^{-k}) \tag{4.25}$$

as $\sigma \rightarrow \infty$ for some $k \geq 2$. Then, with $\tilde{W} = (W - \mathbb{E}W)/\sigma$, and as $\sigma \rightarrow \infty$,

$$d_{\text{loc}}(\mathcal{L}(W), \text{TP}(\mathbb{E}W, \sigma^2)) = O\left(\frac{d_{\text{K}}(\mathcal{L}(\tilde{W}), \Phi)^{1-1/k}}{\sigma}\right)$$

and

$$d_{\text{loc}}(\mathcal{L}(W), \text{TP}(\mathbb{E}W, \sigma^2)) = O\left(\frac{d_{\text{W}}(\mathcal{L}(\tilde{W}), \Phi)^{1-2/(k+1)}}{\sigma}\right).$$

Retaining the previous hypotheses, if (4.25) holds now for some $k \geq 1$, then

$$d_{\text{TV}}(\mathcal{L}(W), \text{TP}(\mathbb{E}W, \sigma^2)) = O(d_{\text{W}}(\mathcal{L}(\tilde{W}), \Phi)^{k/(k+1)}),$$

as $\sigma \rightarrow \infty$.

Proof. First, consider the setup conditional on \mathcal{F} . Divide the sum $Z_1 + \dots + Z_N$ into k successive blocks, each of size $\lfloor N/k \rfloor$, plus one last block with less than $\lfloor N/k \rfloor$ elements. By Lemma 3.4, we have

$$D_k(W|\mathcal{F}) \leq 2 \prod_{l=1}^k \eta_{N,l},$$

where

$$\begin{aligned} \eta_{N,l} &= D_1(Z_{(l-1)\lfloor N/k \rfloor+1} + \dots + Z_{l\lfloor N/k \rfloor}) \\ &\leq \sqrt{\frac{8}{\pi}} \left(\frac{1}{4} + \sum_{i=(l-1)\lfloor N/k \rfloor+1}^{l\lfloor N/k \rfloor} \left(1 - \frac{1}{2} D_1(Z_i) \right) \right)^{-1/2}, \end{aligned}$$

where we have used Lemma 3.5. Therefore, using assumption (4.24), for $l = 1, \dots, k$,

$$\eta_{N,l} \leq \left(\frac{8}{\pi \lfloor N/k \rfloor u} \right)^{1/2}.$$

Now, assume without loss of generality that $\sigma^2 > k/\beta$. In this case

$$\mathbb{I}[N \geq \beta\sigma^2] \frac{1}{\lfloor N/k \rfloor} \leq \frac{1}{\beta\sigma^2/k - 1}$$

and since we can always trivially bound $D_k(W|\mathcal{F})$ by 2^k because $\eta_{N,l} \leq 2$, we have

$$D_k(W|\mathcal{F}) \leq 2^k \mathbb{I}[N < \beta\sigma^2] + \mathbb{I}[N \geq \beta\sigma^2] \left(\frac{8k}{\pi u(\beta\sigma^2 - k)} \right)^{k/2}.$$

Hence, by Lemma 3.3,

$$D_k(W) \leq \mathbb{E}D_k(W|\mathcal{F}) \leq 2^k \mathbb{P}[N < \beta\sigma^2] + \left(\frac{8k}{\pi u(\beta\sigma^2 - k)} \right)^{k/2},$$

which is $O(\sigma^{-k})$ as $\sigma \rightarrow \infty$. After noting that W is integer valued and, hence, $\sigma^{-1} = O(d_W(\mathcal{L}(\tilde{W}), \Phi) \wedge d_K(\mathcal{L}(\tilde{W}), \Phi))$, the claims now follows easily from (ii), (iii), and (iv) of Theorem 2.2, and Lemma 4.1, keeping in mind (4.5). \square

Note that, under the stated conditions, Theorem 4.14 implies the LLT for W if it satisfies the CLT, as the latter also implies convergence in the Kolmogorov metric. If a rate of convergence is available, Theorem 4.14 also yields an upper bound on the rate of convergence for the LLT.

To illustrate Theorem 4.14, we consider the so-called *independence number* of a random graph. The independence number of a graph G is defined to be the maximal number of vertices that can be chosen from the graph so that no two of these vertices are connected.

Consider the following random graph model, which is a simplified version of one discussed by Penrose and Yukich [26]. Let the open set $U \subset \mathbb{R}^d$, $d \geq 1$, be of finite volume, which, without loss of generality, we assume to be 1. Let \mathcal{X} be a homogeneous Poisson point process on U with intensity λ with respect to the Lebesgue measure. Define $G(\mathcal{X}, r)$ to be the graph on the vertex set \mathcal{X} by connecting two vertices whenever they are at most distant r apart from each other. In the context of this random geometric graph, the independence number is the maximal number of closed balls of radius $r/2$ with centers chosen from \mathcal{X} , so that no two balls are intersecting.

Theorem 4.15 (Penrose and Yukich [26]). *For $b > 0$, let W_b be the independence number in $G(\mathcal{X}, b\lambda^{-1/d})$. Then, if b is small enough, we have $\text{Var } W_b \asymp \lambda$ and*

$$d_K \left(\mathcal{L} \left(\frac{W_b - \mathbb{E}W_b}{\sqrt{\text{Var } W_b}} \right), N(0, 1) \right) = O(\log(\lambda)^{3d} \lambda^{-1/2})$$

as $\lambda \rightarrow \infty$.

The condition “ b is small enough” is described in greater detail in Section 2.4 of Penrose and Yukich [26] and is necessary to guarantee the asymptotic order of the variance of W_b . We can give a local limit result as follows.

Theorem 4.16. *Under the assumptions of Theorem 4.15, we have that for every $\varepsilon > 0$,*

$$d_{\text{loc}}(\mathcal{L}(W_b), \text{TP}(\mathbb{E}W_b, \text{Var } W_b)) = O(\lambda^{-1+\varepsilon}),$$

as $\lambda \rightarrow \infty$.

Proof. Let $R = b\lambda^{-1/d}/2$. Denote by $B_R(x)$ the closed ball with radius R and center x ; define $\partial B_R(x) = B_{2R}(x) \setminus B_R(x)$. Now, choose n non-intersecting balls in U , each of radius $3R$ and centers x_1, \dots, x_n ; it is clear that it is possible to have $n \asymp \lambda$. For ball $B_{3R}(x_i)$, define the indicators

$$I_i = \mathbb{I}[\partial B_R(x_i) \cap \mathcal{X} \text{ is empty}], \quad J_i = \mathbb{I}[B_R(x_i) \cap \mathcal{X} \text{ is not empty}].$$

Note that the I_i are independent and identically distributed and, hence, $N = \sum_{i=1}^n I_i \sim \text{Bi}(n, p)$ with $p = \mathbb{E}I_i$ being bounded away from 1 and 0 as $\lambda \rightarrow \infty$. We let $\mathcal{F} = \sigma(I_1, \dots, I_n)$. Furthermore, note that if $I_1 = 1$, then J_i is exactly the contribution of the ball $B_{2R}(x_i)$ to the independence number W_b , as within $B_R(x_i)$ all the vertices are connected and there is no connection to any other vertices outside $B_R(x_i)$. Therefore we can find Y such that

$$W_n = Y + \sum_{j=1}^N J_{K_j},$$

where K_1, \dots, K_N are the indices of those balls with $I_i = 1$. Given \mathcal{F} , note that J_{K_1}, \dots, J_{K_N} are independent $\text{Be}(q)$, with $q = \mathbb{E}J_{K_j}$ being bounded away from 0 and 1, and they are also independent of Y . This implies condition (4.24) for $u = q \wedge (1 - q)$ which is bounded away from 0 as $\lambda \rightarrow \infty$. Using usual exponential tail bounds for the binomial distribution, it is easy to see that, for every k , one can find β such that (4.25) holds. In combination with Theorem 4.15, Theorem 4.14 now yields the claim. \square

Acknowledgements

We are indebted to Steven Evans for pointing out that the inequalities we proved and used in an earlier version of the manuscript were in fact Landau–Kolmogorov inequalities, and thank Peter Eichelsbacher for helpful discussions. We are also grateful to the two anonymous referees for their helpful comments.

Both authors were partially supported by NUS research grant R-155-000-098-133 and NR would like to express his gratitude for the kind hospitality of the Department of Statistics and Applied Probability, NUS, during his research visit. AR was supported by NUS research grant R-155-000-124-112.

References

[1] Barbour, A.D. (2009). Univariate approximations in the infinite occupancy scheme. *ALEA Lat. Am. J. Probab. Math. Stat.* **6** 415–433. MR2576025

- [2] Barbour, A.D. and Čekanavičius, V. (2002). Total variation asymptotics for sums of independent integer random variables. *Ann. Probab.* **30** 509–545. [MR1905850](#)
- [3] Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. New York: The Clarendon Press, Oxford Univ. Press. [MR1163825](#)
- [4] Barbour, A.D., Karoński, M. and Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47** 125–145. [MR1047781](#)
- [5] Barbour, A.D. and Xia, A. (1999). Poisson perturbations. *ESAIM Probab. Statist.* **3** 131–150. [MR1716120](#)
- [6] Behrisch, M., Coja-Oghlan, A. and Kang, M. (2010). The order of the giant component of random hypergraphs. *Random Structures Algorithms* **36** 149–184. [MR2583059](#)
- [7] Bender, E.A., Canfield, E.R. and McKay, B.D. (1997). The asymptotic number of labeled graphs with n vertices, q edges, and no isolated vertices. *J. Combin. Theory Ser. A* **80** 124–150. [MR1472108](#)
- [8] Chatterjee, S. (2007). Stein’s method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. [MR2288072](#)
- [9] Chatterjee, S., Diaconis, P. and Meckes, E. (2005). Exchangeable pairs and Poisson approximation. *Probab. Surv.* **2** 64–106. [MR2121796](#)
- [10] Chatterjee, S. and Shao, Q.-M. (2011). Nonnormal approximation by Stein’s method of exchangeable pairs with application to the Curie–Weiss model. *Ann. Appl. Probab.* **21** 464–483. [MR2807964](#)
- [11] Chen, L.H.Y., Fang, X. and Shao, Q.-M. (2013). From Stein identities to moderate deviations. *Ann. Probab.* **41** 262–293. [MR3059199](#)
- [12] Chen, L.H.Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein’s Method. Probability and Its Applications (New York)*. Heidelberg: Springer. [MR2732624](#)
- [13] Davis, B. and McDonald, D. (1995). An elementary proof of the local central limit theorem. *J. Theoret. Probab.* **8** 693–701. [MR1340834](#)
- [14] Eichelsbacher, P. and Löwe, M. (2010). Stein’s method for dependent random variables occurring in statistical mechanics. *Electron. J. Probab.* **15** 962–988. [MR2659754](#)
- [15] Ellis, R.S. (1985). *Entropy, Large Deviations, and Statistical Mechanics. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **271**. New York: Springer. [MR0793553](#)
- [16] Ellis, R.S., Newman, C.M. and Rosen, J.S. (1980). Limit theorems for sums of dependent random variables occurring in statistical mechanics. II. Conditioning, multiple phases, and metastability. *Z. Wahrsch. Verw. Gebiete* **51** 153–169. [MR0566313](#)
- [17] Gibbs, A.L. and Su, F.E. (2002). On choosing and bounding probability metrics. *Internat. Statist. Rev.* **70** 419–435.
- [18] Goldstein, L. and Xia, A. (2006). Zero biasing and a discrete central limit theorem. *Ann. Probab.* **34** 1782–1806. [MR2271482](#)
- [19] Hardy, G.H., Landau, E. and Littlewood, J.E. (1935). Some inequalities satisfied by the integrals or derivatives of real or analytic functions. *Math. Z.* **39** 677–695. [MR1545530](#)
- [20] Karoński, M. and Łuczak, T. (2002). The phase transition in a random hypergraph. *J. Comput. Appl. Math.* **142** 125–135. [MR1910523](#)
- [21] Kordecki, W. (1990). Normal approximation and isolated vertices in random graphs. In *Random Graphs '87 (Poznań, 1987)* 131–139. Chichester: Wiley. [MR1094128](#)
- [22] Kwong, M.K. and Zettl, A. (1992). *Norm Inequalities for Derivatives and Differences. Lecture Notes in Math.* **1536**. Berlin: Springer. [MR1223546](#)
- [23] Mattner, L. and Roos, B. (2007). A shorter proof of Kanter’s Bessel function concentration bound. *Probab. Theory Related Fields* **139** 191–205. [MR2322695](#)
- [24] McDonald, D.R. (1979). On local limit theorem for integer valued random variables. *Teor. Veroyatn. Primenen.* **24** 607–614. [MR0541375](#)

- [25] Penrose, M.D. and Peres, Y. (2011). Local central limit theorems in stochastic geometry. *Electron. J. Probab.* **16** 2509–2544. [MR2869414](#)
- [26] Penrose, M.D. and Yukich, J.E. (2005). Normal approximation in geometric probability. In *Stein's Method and Applications. Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **5** 37–58. Singapore: Singapore Univ. Press. [MR2201885](#)
- [27] Petrov, V.V. (1975). *Sums of Independent Random Variables*. New York: Springer. Translated from the Russian by A.A. Brown, *Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82*. [MR0388499](#)
- [28] Pósfai, A. (2009). An extension of Mineka's coupling inequality. *Electron. Commun. Probab.* **14** 464–473. [MR2559096](#)
- [29] Röllin, A. (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11** 1115–1128. [MR2189083](#)
- [30] Röllin, A. (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.* **17** 1596–1614. [MR2358635](#)
- [31] Ruciński, A. (1988). When are small subgraphs of a random graph normally distributed? *Probab. Theory Related Fields* **78** 1–10. [MR0940863](#)
- [32] Stepanov, V.E. (1970). The probability of the connectedness of a random graph $\mathcal{G}_m(t)$. *Teor. Veroyatn. Primenen.* **15** 58–68. [MR0270406](#)

Received May 2013 and revised December 2013