Bernoulli **25**(2), 2019, 1076–1104
https://doi.org/10.3150/17-BEJ1013

Error bounds in local limit theorems using Stein's method

A.D. BARBOUR[1], ADRIAN RÖLLIN[2] and NATHAN ROSS[3]

[1]*Institut für Mathematik, Universität Zürich, Winterthurertrasse 190, CH-8057 Zürich, Switzerland. E-mail: a.d.barbour@math.uzh.ch; url: http://user.math.uzh.ch/barbour/*
[2]*Department of Statistics and Applied Probability, 6 Science Drive 2, Singapore, 117546, Singapore. E-mail: adrian.roellin@nus.edu.sg; url: https://blog.nus.edu.sg/roellin/*
[3]*School of Mathematics and Statistics, Peter Hall Building, University of Melbourne, VIC, 3010, Australia. E-mail: nathan.ross@unimelb.edu.au; url: http://www.nathanrossprob.com*

We provide a general result for bounding the difference between point probabilities of integer supported distributions and the translated Poisson distribution, a convenient alternative to the discretized normal. We illustrate our theorem in the context of the Hoeffding combinatorial central limit theorem with integer valued summands, of the number of isolated vertices in an Erdős–Rényi random graph, and of the Curie–Weiss model of magnetism, where we provide optimal or near optimal rates of convergence in the local limit metric. In the Hoeffding example, even the discrete normal approximation bounds seem to be new. The general result follows from Stein's method, and requires a new bound on the Stein solution for the Poisson distribution, which is of general interest.

Keywords: approximation error; Curie–Weiss model; Erdős–Rényi random graph; Hoeffding combinatorial statistic; local limit theorem

1. Introduction

The local limit theorem for general sums W of independent integer valued random variables began with the seminal work of Essen [19], and is now well understood [29], Chapter VII. For sums of dependent random variables, however, much less is known. A key idea, introduced by McDonald [28], is to prove local theorems by using a combination of the corresponding (global) central limit theorem, together with an *a priori* estimate of the smoothness of the distribution $\mathcal{L}(W)$ being approximated. Röllin [30] used this strategy, combined with Stein's method, to develop a systematic approach to approximation by the discrete normal distribution, not only locally, but also globally with respect to the total variation distance.

In both [28] and [30], the smoothness estimates are derived by finding a suitable large collection of conditionally independent Bernoulli random variables embedded in the construction of W. In [34], a fundamentally different technique was discovered, which is instead based on finding a suitable exchangeable pair in the spirit of Stein [36], Chapter I, Lemma 3. They combined it with Landau–Kolmogorov inequalities to give local limit approximations in a variety of examples, but often with less than optimal rates. In this paper, we use Stein's method and the smoothness approach to give a general local limit approximation theorem for settings in which dependence can be described in terms of an (approximate) Stein coupling as given in [13]. This formulation is very flexible, and includes exchangeable pair, local dependence and size-bias settings as particular instances. In the examples that we consider, our approach yields bounds that, when expressed as functions of $\sigma^2 := \operatorname{Var} W$, are no worse than a log factor from the optimal rate of $\mathrm{O}(\sigma^{-2})$. Our general bound is expressed in terms of quantities that typically arise when using Stein's method in the central limit context. As a result, we are able to give bounds for the total variation error in discrete normal approximation as well as the local limit bounds with no extra effort.

1.1. Translated Poisson distribution

As in [31], we use *translated Poisson* distributions as approximating family instead of discretised normal distributions – Lemma 1.1 justifies this to the accuracy of interest to us. We say that the random variable Z has the translated Poisson distribution and write $Z \sim \mathrm{TP}(\mu, \sigma^2)$ if $Z - s \sim \mathcal{P}(\sigma^2 + \gamma)$, where

$$s := \lfloor \mu - \sigma^2 \rfloor, \qquad \gamma := \mu - \sigma^2 - \lfloor \mu - \sigma^2 \rfloor, \tag{1.1}$$

and where $\mathcal{P}(\lambda)$ denotes the Poisson distribution with mean λ. Note that $\mathbb{E}Z = \mu$ and $\sigma^2 \le \operatorname{Var} Z \le \sigma^2 + 1$. The translated Poisson distribution is a Poisson distribution, but translated by an integer chosen so that both its mean and variance closely match prescribed values μ and σ^2. The following lemma say that the translated Poisson distribution is an appropriate substitute for the discretized normal distribution. Its proof follows easily from the classical local central limit theorem with error.

Lemma 1.1. *There exists a constant $C > 0$ such that, for all $\mu \in \mathbb{R}$ and $\sigma^2 \ge 1$,*

$$\sup_{n\in\mathbb{Z}} \left| \mathrm{TP}\big(\mu, \sigma^2\big)\{n\} - \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(n-\mu)^2}{2\sigma^2}\right) \right| \le \frac{C}{\sigma^2}.$$

We also note some basic properties of the translated Poisson distributions. Define the following "smoothness" measure of an integer valued distribution,

$$S_l\big(\mathcal{L}(W)\big) := \sup_{h:\|h\|\le 1} \left|\mathbb{E}\Delta^l h(W)\right|, \qquad l \ge 1, \tag{1.2}$$

where Δ denotes the first difference operator $\Delta g(k) := g(k+1) - g(k)$. Variations of the smoothing terms (1.2) frequently appear in integer supported distributional approximation results; see, for example, [5,24,32] and [20].

The next result shows the typical smoothness expected for approximately discretized normal distributions. It is shown in [34], Lemma 4.1.

Lemma 1.2. *For each $k \ge 1$ there exists a constant $C(k)$ such that, for all $\mu \in \mathbb{R}$ and $\sigma^2 \ge 1$,*

$$S_k\big(\mathrm{TP}\big(\mu, \sigma^2\big)\big) \le \frac{C(k)}{\sigma^k}. \tag{1.3}$$

1.2. Stein couplings

Our approximations are designed for random variables W that form part of a *Stein coupling*. Following Chen and Röllin [13], we say that the random variables (W, W', G, R) with $\mathbb{E}W = \mu$ form an *approximate Stein coupling* if

$$\mathbb{E}\big[G\big(f\big(W'\big) - f(W)\big)\big] = \mathbb{E}\big[(W-\mu)f(W)\big] + \mathbb{E}\big[Rf(W)\big] \tag{1.4}$$

for all f such that the expectations exist. If $R = 0$ almost surely, we call (W, W', G) a Stein coupling. Some examples of Stein couplings well used in Stein's method are the following:

Local dependence. Let $W = \sum_{i=1}^n X_i$, with $\mathbb{E}X_i = \mu_i$ for $1 \le i \le n$. Suppose that, for each i, there is $A_i \subset \{1, \dots, n\}$ such that X_i is independent of $(X_j)_{j \notin A_i}$. Then, for I a random index, uniformly distributed on $\{1, \dots, n\}$ and independent of $(X_i)_{i=1}^n$,

$$\big(W, W', G\big) := \Big(W, W - \sum_{j \in A_I} X_j, -n(X_I - \mu_I)\Big) \tag{1.5}$$

is a Stein coupling.

Size bias. If W^s has the size bias distribution of W and $\mathbb{E}W = \mu$, then

$$\big(W, W', G\big) := \big(W, W^s, \mu\big)$$

is a Stein coupling.

Exchangeable pairs. If (W, W') is an exchangeable pair satisfying the linearity condition

$$\mathbb{E}\big[W' - W | W\big] = -a(W - \mu) + aR, \tag{1.6}$$

then

$$\big(W, W', G, R\big) := \Big(W, W', \frac{W' - W}{2a}, R\Big)$$

is an approximate Stein coupling.

Exchangeable pairs, one-sided version. If (W, W') is an exchangeable pair that satisfies (1.6), then

$$\big(W, W', G, R\big) := \Big(W, W', \frac{W' - W}{a} \mathrm{I}\big[W' - W > 0\big], R\Big) \tag{1.7}$$

is an approximate Stein coupling.

Note also that, for (W, W', G, R) an approximate Stein coupling,

$$\mathbb{E}\big[G\big(W' - W\big)\big] = \sigma^2 + \mathbb{E}\big[R(W - \mu)\big], \tag{1.8}$$

which can be seen by taking $f(x) = x$ and $f(x) = 1$ (to find $\mathbb{E}R = 0$) in the defining relation (1.4). In particular, if $R = 0$ almost surely, then $\mathbb{E}[G(W' - W)] = \sigma^2$.

2. Main results and applications

We bound the error in the approximation by the translated Poisson distribution of the distributions $\mathcal{L}(W)$ of integer valued random variables with finite variances that can be represented as the W in an (approximate) Stein coupling. Our bounds are expressed in terms of the moments of W and of expectations involving the quantities G and $D := W' - W$, and the conditional smoothness coefficients $S_l(\mathcal{L}(W|\mathcal{F}))$ for some appropriate associated sigma-field $\mathcal{F}$. Exchangeable pairs, size-biasing, and local dependence appear ubiquitously when using Stein's method for distributional approximation and concentration inequalities, so (1) many of the terms appearing in our bound can be fruitfully bounded using well-established techniques, and (2) new techniques developed here for bounding commonly appearing terms will prove useful in other applications of Stein's method.

2.1. An abstract bound

In order to express the accuracy of translated Poisson approximation, we define the *total variation metric* as

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(X), \mathcal{L}(Y)\bigr) := \sup_{A \subseteq \mathbb{Z}} \bigl|\mathbb{P}[X \in A] - \mathbb{P}[Y \in A]\bigr|,$$

as well as a metric to capture the local differences as

$$d_{\mathrm{loc}}\bigl(\mathcal{L}(X), \mathcal{L}(Y)\bigr) := \sup_{a \in \mathbb{Z}} \bigl|\mathbb{P}[X = a] - \mathbb{P}[Y = a]\bigr|.$$

We can now state our main general approximation result, which is proved in Section 3.

Theorem 2.1. *Let (W, W', G, R) be an approximate Stein coupling with W and W' integer valued, $\mathbb{E}W = \mu$ and $\mathrm{Var}(W) = \sigma^2$. Set $D := W' - W$, and let $\mathcal{F}_1$ and $\mathcal{F}_2$ be sigma-algebras such that W is $\mathcal{F}_1$-measurable and such that (G, D) is $\mathcal{F}_2$-measurable. Define*

$$\Psi := \bigl|\mathbb{E}[GD|\mathcal{F}_1] - \mathbb{E}[GD]\bigr|, \qquad \Upsilon := \mathbb{E}\bigl[\bigl|GD(D-1)\bigr|S_2\bigl(\mathcal{L}(W|\mathcal{F}_2)\bigr)\bigr].$$

Then

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathrm{TP}\bigl(\mu, \sigma^2\bigr)\bigr) \le \frac{\mathbb{E}\Psi}{\sigma^2} + \frac{2\sqrt{\mathbb{E}R^2}}{\sigma} + \frac{2(\Upsilon + 1)}{\sigma}, \tag{2.1}$$

and

$$\begin{aligned} d_{\mathrm{loc}}\bigl(\mathcal{L}(W), \mathrm{TP}\bigl(\mu, \sigma^2\bigr)\bigr) \le{}& \frac{\mathbb{E}\Psi}{\sigma^3\sqrt{2e}} + \frac{\mathbb{E}[\Psi|W-\mu|]}{\sigma^4} + \sup_{a\in\mathbb{Z}} \frac{\mathbb{E}\{\Psi \mathrm{I}[W=a]\}}{\sigma^2} \\ &+ \frac{\sqrt{\mathbb{E}R^2}}{\sigma^2}\left(2 + \frac{1}{\sqrt{2e}} + \sigma \sup_{a\in\mathbb{Z}} \mathbb{P}(W=a)\right) + \frac{2(\Upsilon+1)}{\sigma^2}. \end{aligned} \tag{2.2}$$

Note that we have distinguished Ψ and Υ as the significant quantities in the bound, but that Ψ is random while Υ is not.

Remark 2.2. If (W, W') is an exchangeable pair such that (1.6) holds, and if we assume in addition that $D \in \{-1, 0, +1\}$, then we can use the coupling (1.7); in this case, Ψ and Υ simplify to

$$\Psi = \frac{1}{a}\bigl|\mathbb{P}(D=1|\mathcal{F}_1) - \mathbb{P}(D=1)\bigr| \quad \text{and} \quad \Upsilon = 0. \tag{2.3}$$

2.2. Towards a concrete bound

The bounds in Theorem 2.1 are still rather abstract, and it may not be obvious how to handle the individual terms in concrete applications. We now show that, by making a natural additional assumption, the terms appearing in (2.1) and (2.2) can be made more manageable.

To this end, we assume that, for some $\kappa > 0$, for some integer $k \geq 0$ and for some non-negative random variable T, we have

$$\Psi \leq \sigma\kappa \sum_{j=0}^{k} \left(\frac{|W-\mu|}{\sigma}\right)^j + T. \tag{2.4}$$

This assumption appears naturally in many applications, and it is worthwhile emphasising that it is weaker than similar conditions appearing in the literature around Stein's method, such as Condition (3.3) in Theorem 3.1 of Chen, Fang and Shao [12] or the condition in Theorem 3.11 of Röllin [31]. We now have the following easy corollary of Theorem 2.1.

Corollary 2.3. *Under the conditions of Theorem 2.1 and assuming in addition (2.4), we have*

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2)\bigr) \leq \frac{\kappa}{\sigma} \sum_{j=0}^{k} \frac{\mathbb{E}|W-\mu|^j}{\sigma^j} + \frac{\mathbb{E}T}{\sigma^2} + \frac{2\sqrt{\mathbb{E}R^2}}{\sigma} + \frac{2(\Upsilon+1)}{\sigma} \tag{2.5}$$

and

$$\begin{aligned} & d_{\mathrm{loc}}\bigl(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2)\bigr) \\ &\quad \leq \frac{2\kappa}{\sigma^2} \sum_{j=0}^{k+1} \frac{\mathbb{E}|W-\mu|^j}{\sigma^j} + \frac{\kappa}{\sigma^2} \sup_{a\in\mathbb{Z}} \left(\mathbb{P}(W=a) \sum_{j=0}^{k} \frac{|a-\mu|^j}{\sigma^{j-1}}\right) && (2.6) \\ &\qquad + \frac{2\sqrt{\mathbb{E}T^2}}{\sigma^3} + \frac{\sup_{a\in\mathbb{Z}} \mathbb{E}[T\,\mathrm{I}[W=a]]}{\sigma^2} && (2.7) \\ &\qquad + \frac{\sqrt{\mathbb{E}R^2}}{\sigma^2}\Bigl(3 + \sigma \sup_{a\in\mathbb{Z}} \mathbb{P}(W=a)\Bigr) && (2.8) \\ &\qquad + \frac{2(\Upsilon+1)}{\sigma^2}. && (2.9) \end{aligned}$$

Assumption (2.4) is always satisfied by taking $T = \Psi$ and an empty sum, so its real use is if T is more easily managed than Ψ – for instance, if $T = 0$ almost surely. In what follows, we take assumption (2.4) to be satisfied, and consider the bounds (2.6)–(2.9) in turn. We tacitly assume throughout the following discussion that we have a sequence of integer valued random variables $W = W_m$ with means $\mu = \mu_m$ and whose variances $\sigma^2 = \sigma_m^2$ grow to infinity with m; order estimates are to be understood as $m \to \infty$, and the dependence on m is suppressed in the notation.

First, we expect $\mathbb{E}\{|\sigma^{-1}(W-\mu)|^{k+1}\}$ to be bounded, so that the first term of (2.6) is of our target order $\mathrm{O}(\sigma^{-2})$. For the second term of (2.6), we have the following lemma.

Lemma 2.4. *Write $\delta := d_{\mathrm{TV}}(\mathcal{L}(W), \mathrm{TP}(\mu, \sigma^2))$ and assume that for some $1/2 < \alpha \le 1$, we have*

$$\delta = \mathrm{O}(\sigma^{-\alpha}) \quad \textit{and} \quad S_2(\mathcal{L}(W)) = \mathrm{O}(\sigma^{-1-\alpha}).$$

Then, for any $\ell \ge j \ge 1$,

$$\sup_{a\in\mathbb{Z}} \mathbb{P}(W=a)\frac{|a-\mu|^j}{\sigma^{j-1}} \le \mathrm{O}(1) + \mathbb{E}\left\{\left(\frac{|W-\mu|}{\sigma}\right)^{\ell}\right\}\sigma^{1/2+\alpha-(2\alpha-1)\ell/(2j)}$$

and

$$\sup_{a\in\mathbb{Z}} \mathbb{P}(W=a) = \mathrm{O}(\sigma^{-1}).$$

Proof. Recall the definitions $s = \lfloor \mu - \sigma^2 \rfloor$ and $\gamma = \mu - \sigma^2 - s$, and let $\mathcal{P}_\lambda(\cdot) := \mathcal{P}(\lambda)\{\cdot\}$. First note that

$$\sup_{a\in\mathbb{Z}} \mathbb{P}(W=a) \le \delta + \mathrm{O}(\sigma^{-1}),$$

which follows easily from the definition of total variation, and because $\sup_{a\in\mathbb{Z}} \mathcal{P}_\lambda(a) = \mathrm{O}(\lambda^{-1/2})$ as $\lambda \to \infty$, where $\lambda = \sigma^2 + \gamma$.

For the first assertion, we first bound $\mathbb{P}(W=a)$ for a "near" μ. Note that the second assertion follows from (2.10) below, used in this argument, and the last sentence of the previous paragraph. Now, [34], Theorem 2.2(i) with $l=2$, $m=1$, and Lemma 3.1, implies that for some constant C,

$$d_{\mathrm{loc}}\big(\mathcal{L}(W), \mathrm{TP}(\mu,\sigma^2)\big) \le C d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathrm{TP}(\mu,\sigma^s)\big)^{1/2}\big(S_2(\mathcal{L}(W)) + S_2(\mathrm{TP}(\mu,\sigma^2))\big)^{1/2},$$

which, with (1.3) and the hypotheses of the lemma, implies

$$d_{\mathrm{loc}}\big(\mathcal{L}(W), \mathrm{TP}(\mu,\sigma^2)\big) = \mathrm{O}(\sigma^{-1/2-\alpha}).$$

Hence

$$\big|\mathbb{P}(W=a) - \mathcal{P}_{\sigma^2+\gamma}(a-s)\big| = \mathrm{O}(\sigma^{-1/2-\alpha}), \tag{2.10}$$

so that

$$\mathbb{P}(W=a)\frac{|a-\mu|^j}{\sigma^{j-1}} \le \frac{|a-\mu|^j}{\sigma^{j-1/2+\alpha}} + \mathcal{P}_{\sigma^2+\gamma}(a-s)\frac{|a-\mu|^j}{\sigma^{j-1}}.$$

Combining this inequality with the observation that

$$\sup_{\lambda\geq 1}\sup_{r\in\mathbb{Z}} \mathcal{P}_\lambda(r)\frac{|r-\lambda|^j}{\lambda^{(j-1)/2}} < \infty,$$

and noting that, for $r = a - s$ and $\lambda = \sigma^2 + \gamma$,

$$r - \lambda = a - \left(\mu - \sigma^2 - \gamma\right) - \left(\sigma^2 + \gamma\right) = a - \mu,$$

it follows that $\mathbb{P}(W = a)\{\sigma^{-(j-1)}|a-\mu|^j\} = \mathrm{O}(1)$ for $\sigma^{-1}|a-\mu| \leq \sigma^{(2\alpha-1)/(2j)}$.

For values of a "far" from μ, that is, $\sigma^{-1}|a-\mu| > \sigma^{(2\alpha-1)/(2j)}$, use Markov's inequality to give

$$\begin{aligned}\frac{|a-\mu|^j}{\sigma^{j-1}}\mathbb{P}(W=a) &\leq \frac{\sigma|a-\mu|^j}{\sigma^j}\mathbb{P}\left(\frac{|W-\mu|}{\sigma} \geq \frac{|a-\mu|}{\sigma}\right)\\ &\leq \sigma\left|\frac{a-\mu}{\sigma}\right|^{-(l-j)}\mathbb{E}\left\{\left(\frac{|W-\mu|}{\sigma}\right)^\ell\right\}\\ &\leq \sigma\mathbb{E}\left\{\left(\frac{|W-\mu|}{\sigma}\right)^\ell\right\}\sigma^{(2\alpha-1)(j-\ell)/2j},\end{aligned}$$

concluding the proof. □

Remark 2.5. Thus, if (2.4) and the hypotheses of Lemma 2.4 are satisfied, and if $\mathbb{E}|W-\mu|^K = \mathrm{O}(\sigma^K)$ for some $K \geq k(1+2\alpha)/(2\alpha-1)$, with α as in Lemma 2.4, then the second term of (2.6) is of order $\mathrm{O}(\sigma^{-2})$.

We next show that, if T is concentrated around zero, then the two terms of (2.7) can be suitably bounded.

Lemma 2.6. *Suppose that the nonnegative random variable $T = T_m$ satisfies*

$$\mathbb{E}\left(\sigma^{-1}T\mathrm{I}\left[\sigma^{-1}T \geq t\right]\right) \leq \varepsilon(t), \qquad t \geq 1 \tag{2.11}$$

for some $\varepsilon(t)$ with $\int_1^\infty \varepsilon(t)\,dt < K < \infty$, and K is the same for all m. Then $\mathbb{E}T^2 = \mathrm{O}(\sigma^2)$.

If (2.11) is satisfied, then for any $k \in \mathbb{Z}$ and $t \geq 1$,

$$\mathbb{E}\left[T\mathrm{I}[W=k]\right] \leq \sigma\varepsilon(t) + t\sigma\sup_{a\in\mathbb{Z}}\mathbb{P}(W=a).$$

Proof. By a standard calculation, $\mathbb{E}T^2 \leq \sigma^2(1+\int_1^\infty \varepsilon(t)\,dt)$. For the second assertion, note that

$$\mathbb{E}\left[T\mathrm{I}[W=k]\right] \leq \mathbb{E}\left\{T\mathrm{I}[T\geq t\sigma]\right\} + t\sigma\mathbb{P}(W=k).$$

The former term is bounded by $\sigma\varepsilon(t)$ and the latter by $t\sigma\sup_{a\in\mathbb{Z}}\mathbb{P}(W=a)$. □

Remark 2.7. For example, suppose that $\sup_{a\in\mathbb{Z}} \mathbb{P}(W=a) = \mathrm{O}(\sigma^{-1})$. Then if $\varepsilon(t)=0$ for all $t \geq t_0$, for some $t_0 < \infty$, we have a bound for (2.7) of the ideal order $\mathrm{O}(\sigma^{-2})$. If, for some constant c, we have $\varepsilon(t) \leq e^{-t^2/2c}$, then the choice $t = \sqrt{2c\log\sigma}$ gives $\sup_k \mathbb{E}[T\mathrm{I}[W=k]] = \mathrm{O}(\sqrt{\log\sigma})$, and a bound for (2.7) of order $\mathrm{O}(\sigma^{-2}\sqrt{\log\sigma})$. If $\varepsilon(t) \leq e^{-t/c}$, then the choice $t = c\log\sigma$ gives a bound for (2.7) of order $\mathrm{O}(\sigma^{-2}\log\sigma)$. Note also that, under the conditions of Lemma 2.4,

$$\mathbb{E}T\mathbb{P}[W=k] \leq \sqrt{\mathbb{E}T^2}\mathbb{P}[W=k] = \mathrm{O}(1),$$

and so, in (2.7), $\sup_a \mathbb{E}[T\mathrm{I}[W=a]]$ can be replaced by

$$\sup_a \bigl|\mathrm{Cov}\bigl(T, \mathrm{I}[W=a]\bigr)\bigr| + \mathrm{O}(1).$$

For the remaining terms in Corollary 2.3, if $\mathbb{E}R^2 = \mathrm{O}(1)$, then it is easy to see that (2.8) is of order $\mathrm{O}(\sigma^{-1}\sup_{a\in\mathbb{Z}} \mathbb{P}(W=a) + \sigma^{-2})$. This leaves Υ, which is handled using the methods discussed in [34] and illustrated in the applications below. We collect the results above in the following corollary.

Corollary 2.8. *Assume the notation and hypotheses of Theorem 2.1 and suppose that (2.4) is satisfied for some choice of κ, k and T.*

(i) *If $\mathbb{E}\{(\sigma^{-1}|W-\mu|)^k\} = \mathrm{O}(1)$, $\mathbb{E}T = \mathrm{O}(\sigma)$ and $\mathbb{E}R^2 = \mathrm{O}(1)$, then it follows that*

$$\delta := d_{\mathrm{TV}}\bigl(\mathcal{L}(W), \mathrm{TP}\bigl(\mu,\sigma^2\bigr)\bigr) = \mathrm{O}\bigl(\sigma^{-1}(1+\Upsilon)\bigr).$$

(ii) *If $\mathbb{E}R^2 = \mathrm{O}(1)$ and, for some $1/2 < \alpha \leq 1$,*

(1) *$\delta = \mathrm{O}(\sigma^{-\alpha})$ and $S_2(\mathcal{L}(W)) = \mathrm{O}(\sigma^{-1-\alpha})$;*

(2) *$\mathbb{E}\{(\sigma^{-1}|W-\mu|)^K\} = \mathrm{O}(1)$ for some $K \geq k(1+2\alpha)/(2\alpha-1)$;*

(3) *T/σ is almost surely uniformly bounded,*

then

$$d_{\mathrm{loc}}\bigl(\mathcal{L}(W), \mathrm{TP}\bigl(\mu,\sigma^2\bigr)\bigr) = \mathrm{O}\bigl(\sigma^{-2}(\Upsilon+1)\bigr). \tag{2.12}$$

If (3) is replaced by

(3a) *$\sup_a|\mathrm{Cov}(T, \mathrm{I}[W=a])|$ is bounded,*

then (2.12) still holds. If (3) is replaced by

(3b) *$\varepsilon(t)$ of Lemma 2.6 has an exponential tail,*

then the term $\mathrm{O}(\sigma^{-2}(\Upsilon+1))$ has to be replaced by $\mathrm{O}(\sigma^{-2}(\Upsilon+1)\log\sigma)$ in the bound in (2.12).

Note that the value of α used in (1) and (2) does not appear in the error estimate; the assumptions are there to ensure that enough moments of $\sigma^{-1}|W-\mu|$ are finite. However, if the largest α for which $S_2(\mathcal{L}(W)) = \mathrm{O}(\sigma^{-1-\alpha})$ is such that $\alpha < 1$, then, even in the ideal case in which $|GD(D-1)| = \mathrm{O}(\sigma^2)$ almost surely, the quantity Υ is only guaranteed to be of order $\mathrm{O}(\sigma^{1-\alpha})$, yielding a bound in (2.12) of order $\mathrm{O}(\sigma^{-1-\alpha})$, and not of the ideal order $\mathrm{O}(\sigma^{-2})$.

Remark 2.9 (Sums of independent random variables). If $W_n = \sum_{i=1}^n X_i$, where the X_i are independent integer valued random variables such that $\sum_{i=1}^n \mathbb{E}|X_i - \mathbb{E}X_i|^3 = \mathrm{O}(\sigma_n^2)$, and which satisfy an aperiodicity condition, then Theorems 4 and 5 in Chapter VII of [29] imply that the error made in the local limit approximation by the discrete normal (and hence the translated Poisson) is of best order σ_n^{-2}. If we assume the somewhat stronger aperiodicity assumption, that $\sigma_n^{-2}\sum_{i=1}^n(1 - d_{\mathrm{TV}}(\mathcal{L}(X_i), \mathcal{L}(X_i + 1)))$ is bounded away from zero, then it follows that $S_2(\mathcal{L}(W_n)) = \mathrm{O}(\sigma_n^{-2})$ and that Υ is also of the correct order for good rates, so that the only problem term, in the decomposition (2.4), is $\sup_{k\in\mathbb{Z}} \mathbb{E}[T\mathrm{I}[W = k]]$. Using our approach, in conjunction with the local dependence Stein coupling at (1.5) with $A_i = \{i\}$, we deduce a $T = \Psi$ of the form

$$\left|\sum_{i=1}^n X_i(X_i - \mu_i) - \sigma_n^2\right|,$$

and, as in Corollary 2.8, this together with Lemma 2.6 leads to bounds that depend strongly on the tail behaviour of X_i. For example, if the X_i have finite $(2j)$th moment for some $j \geq 2$, then by Hölder's and Rosenthal's inequalities,

$$\sup_{k\in\mathbb{Z}} \mathbb{E}\big[T\mathrm{I}[W = k]\big] \leq \big(\mathbb{E}T^j\big)^{1/j} \sup_{k\in\mathbb{Z}}\big(\mathbb{P}(W = k)\big)^{(j-1)/j} = \mathrm{O}\big(n^{-1/(2j)}\big),$$

with a constant depending on j, implying an upper bound on the local metric of sub-optimal order $n^{-1+1/(2j)}$. Thus a direct application of our approach, which can be effective in much more challenging applications, is sub-optimal in this classical case. As it happens, a small modification of the proof of Lemma 3.5 below, adding and subtracting $GD\Delta f(W')$ rather than $GD\Delta f(W)$ after (3.8), eliminates the problem term, and leads to an approximation error of the same asymptotic order as that given in [29], Theorems 4 and 5 in Chapter VII, albeit under our stronger aperiodicity assumption. This is essentially the approach taken by Röllin [32], Theorem 2.1.

2.3. Hoeffding permutation statistic

Let $(a_{ij})_{1\leq i,j\leq n}$ be an array of integers, and define

$$W := \sum_{i=1}^n a_{i\rho_i},$$

where ρ is a uniformly chosen random permutation. Defining

$$a_{i+} := \sum_{j=1}^n a_{ij}, \qquad a_{+j} := \sum_{i=1}^n a_{ij}, \qquad a_{++} := \sum_{i,j=1}^n a_{ij},$$
$$\hat{a}_{ij} := a_{ij} - \frac{a_{i+}}{n} - \frac{a_{+j}}{n} + \frac{a_{++}}{n^2},$$

we have

$$\mu := \mathbb{E}W = \frac{1}{n}a_{++} \quad \text{and} \quad \sigma^2 := \operatorname{Var} W = \frac{1}{n-1}\sum_{i,j}\hat{a}_{ij}^2. \tag{2.13}$$

We are interested in the accuracy of local approximation to $\mathcal{L}(W)$ by $\mathrm{TP}(\mu, \sigma^2)$.

Central limit theorems for W have a long history going back to [37] and [25]. More recent refinements obtaining Berry–Esseen error bounds under various conditions on the matrix a were derived by [7,11,21]; see references of the last for an up to date history.

Our main results are in terms of asymptotic rates as $n \to \infty$ for a sequence of such matrices $a^{(n)}$, assuming that, for suitable positive constants A_1, α_0, α_1, and α_2,

- *Assumption A1.* $\max_{1\le i,j\le n} |a_{ij}^{(n)}| \le A_1 < \infty$ and $n^{-1}(\sigma^{(n)})^2 \ge (\alpha_0 A_1)^2 > 0$ for all n.
- *Assumption A2.* There exists a set $\mathcal{I} := \{\{i_{l1}, i_{l2}\}, 1 \le l \le n_1\}$ of $n_1 \ge \alpha_1 n$ disjoint pairs of indices in $[n]$ such that, for $\{i_1, i_2\} \in \mathcal{I}$, there exists a set $\mathcal{J}(i_1, i_2)$ of $n_2 \ge \alpha_2 n^2$ pairs (clearly *not* disjoint) of indices $\{j_1, j_2\}$ such that

$$|a_{i_1,j_1} + a_{i_2,j_2} - a_{i_1,j_2} - a_{i_2,j_1}| = 1.$$

We also assume without loss that $|\mu| \le n/2$, by replacing a_{ij} by $a_{ij} + m$ for all i, j, for a suitably chosen integer m. Assumption A1 is a standard simplifying assumption when studying the Hoeffding permutation statistic and something like Assumption A2 is necessary to ensure that W is not concentrated on a sub-lattice of $\mathbb{Z}$. For instance, if all the a_{ij} are even, the distribution of W lies on the lattice $2\mathbb{Z}$, and then $S_2(\mathcal{L}(W)) = 4$; so some additional conditions on the matrix a are needed to ensure smoothness. Our methods still apply if either of these assumptions are weakened, but at the cost of worse bounds or greater technicality.

Our main result is as follows.

Theorem 2.10. *Let $a^{(n)}$ be a sequence of matrices satisfying Assumptions* A1 *and* A2 *and let $W = W_n$ be the Hoeffding permutation statistic defined above. Then,*

$$d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathrm{TP}\big(\mathbb{E}W, \operatorname{Var}(W)\big)\big) = \mathrm{O}\big(\sigma^{-1}\big),$$

$$d_{\mathrm{loc}}\big(\mathcal{L}(W), \mathrm{TP}\big(\mathbb{E}W, \operatorname{Var}(W)\big)\big) = \mathrm{O}\bigg(\frac{\sqrt{\log(\sigma)}}{\sigma^2}\bigg).$$

2.4. Isolated vertices in the Erdős–Rényi random graph

We show a local limit bound with optimal rate for W defined to be the number of isolated vertices in an Erdős–Rényi graph on n vertices with edge probability $p \sim \lambda/n$ for some $\lambda > 0$. Note that

$$\mu := \mathbb{E}W = n(1-p)^{n-1}; \qquad \sigma^2 := \operatorname{Var}(W) = n(1-p)^{n-1}\big[1 + (np-1)(1-p)^{n-2}\big],$$

and so in the regime $p \sim \lambda/n$, we have $\mu \sim ne^{-\lambda}$ and $\sigma^2 \sim ne^{-\lambda}\{1 + (\lambda - 1)e^{-\lambda}\}$ are of strict order n.

Studying degree and subgraph count statistics to understand the structure of Erdős–Rényi graphs has a long history, and is still an active area to this day; for example, [27] and [33]. A number of works derive central limit theorems with error rates for isolated degrees. Error rates for smooth test function metrics are provided by Barbour, Karoński and Ruciński [4] and Kordecki [26]; for Kolmogorov distance by [22]; and for total variation distance (to a discretized normal) by [20]. We show the following optimal local limit theorem that strengthens the rate provided in [34].

Theorem 2.11. *Let $W = W_n$ be the number of isolated vertices in an Erdős–Rényi graph on n vertices with edge probability $p \sim \lambda/n$. Then*

$$d_{\mathrm{loc}}\big(\mathcal{L}(W), \mathrm{TP}\big(\mathbb{E}W, \mathrm{Var}(W)\big)\big) = \mathrm{O}\bigg(\frac{\sqrt{\log(\sigma)}}{\sigma^2}\bigg).$$

2.5. Magnetization in the Curie–Weiss model

The Curie–Weiss model on n sites is given by a Gibbs measure on $\{-1, +1\}^n$ having parameters $\beta > 0$ and $h \in \mathbb{R}$. The random vector $S = (S_1, \dots, S_n) \in \{-1, +1\}^n$ has this distribution if

$$\mathbb{P}\big(S = (s_1, \dots, s_n)\big) = Z_{\beta,h}^{-1} \exp\Bigg\{\frac{\beta}{n} \sum_{1 \le i < j \le n} s_i s_j + h \sum_{i=1}^{n} s_i\Bigg\}. \tag{2.14}$$

The *magnetization* $W = \sum_{i=1}^n S_i$ of the system has been the object of intense study over the last forty years or more; see [16], Section IV.4. By symmetry, we only need consider $h \ge 0$.

We first state the law of large numbers for W/n, which relies on the following equation for fixed $\beta > 0$, $h \ge 0$:

$$m = \tanh(\beta m + h). \tag{2.15}$$

For $h > 0$, there is only one *positive* solution m_h satisfying (2.15). If $h = 0$ and $0 < \beta < 1$, $m_0 = 0$ is the only solution to (2.15).

Lemma 2.12 ([16], Theorem IV.4.1). *If S is distributed as (2.14) for some $h > 0$ and $\beta > 0$ or for $h = 0$ and $0 < \beta < 1$, and if $W = \sum_{i=1}^n S_i$, then as $n \to \infty$,*

$$\frac{W}{n} \xrightarrow{\text{prob}} m_h.$$

We then have the following distributional convergence result from [17], Theorem 2.1; see also [18].

Theorem 2.13. *If S is distributed as (2.14) for some $h > 0$ and $\beta > 0$ or for $h = 0$ and $0 < \beta < 1$, and if $W = \sum_{i=1}^n S_i$, then as $n \to \infty$,*

$$\mathcal{L}\bigg(\frac{W - nm_h}{\sqrt{n}}\bigg) \to \mathrm{N}\bigg(0, \frac{1 - m_h^2}{1 - \beta + \beta m_h^2}\bigg).$$

Above the critical temperature, a convergence rate of order $\mathrm{O}(n^{-1/2})$ in Kolmogorov distance is a consequence of Barbour [2], Theorem 3 and pp. 602–605; see also [10] and [15]. Concentration inequalities are derived in [8] and [9]; total variation and local limit bounds (that are weaker than those obtained below) are given in [34]. Note also that, for $\mu_n := \mathbb{E}W$ and $\sigma_n^2 := \mathrm{Var}(W)$, $\mu_n \sim nm_h$ and $\sigma_n^2 \sim n(1-m_h^2)/(1-\beta+\beta m_h^2)$ as $n\to\infty$.

Our main result for the magnetization is a sharp rate of convergence in the local limit metric. Note that W sits on a lattice of span 2, so we ultimately shift and scale to put it on $\{0,\dots,n\}$.

Theorem 2.14. *Let S be distributed as (2.14) for some $h>0$ and $\beta>0$ or for $h=0$ and $0<\beta<1$, $W=W_n=\sum_{i=1}^n S_i$, and $\widetilde{W} := (W+\frac{1}{2}\{1-(-1)^n\})/2$. Then in the notation above,*

$$d_{\mathrm{TV}}\biggl(\mathcal{L}(\widetilde{W}),\mathrm{TP}\biggl(\frac{nm_h}{2},\frac{n(1-m_h^2)}{4(1-\beta+\beta m_h^2)}\biggr)\biggr) = \mathrm{O}\bigl(\sigma_n^{-1}\bigr) = \mathrm{O}\bigl(n^{-1/2}\bigr),$$

$$d_{\mathrm{loc}}\biggl(\mathcal{L}(\widetilde{W}),\mathrm{TP}\biggl(\frac{nm_h}{2},\frac{n(1-m_h^2)}{4(1-\beta+\beta m_h^2)}\biggr)\biggr) = \mathrm{O}\bigl(\sigma_n^{-2}\bigr) = \mathrm{O}\bigl(n^{-1}\bigr).$$

The remainder of the paper is devoted to proofs of the results above. Theorem 2.1 is proved in the next section, and application statements are proved in Section 4.

3. Proof of Theorem 2.1

3.1. Preliminaries

To express the accuracy of approximation by a translated Poisson distribution using Stein's method, we need the solutions $(g_A)_{A\subset\mathbb{Z}^+}$ of the Poisson Stein equation

$$\lambda\Delta g_A(i) - (i-\lambda)g_A(i) = \mathrm{I}[i\in A] - \mathcal{P}(\lambda)\{A\}, \qquad i\ge 0; \qquad g_A(i)=0, \qquad i\le 0. \quad (3.1)$$

For approximation by $\mathrm{TP}(\mu,\sigma^2)$, defining s and γ as in (1.1), we take $\lambda := \sigma^2+\gamma$ and define $f_A:\mathbb{Z}\to\mathbb{R}$ by $f_A(i) := g_A(i-s)$. Where there is no likelihood of confusion, we write f_a for $f_{\{a\}}$. Then the following representations of the accuracy of translated Poisson approximation to an integer valued random variable W were shown in [31], (3.18).

Proposition 3.1. *Let W be an integer valued random variable with mean μ and variance σ^2, and let f_A be defined as above. Then*

$$d_{\mathrm{TV}}\bigl(\mathcal{L}(W),\mathrm{TP}\bigl(\mu,\sigma^2\bigr)\bigr) \le \sup_{A\subseteq\mathbb{Z}^+}\bigl|\mathbb{E}\sigma^2\Delta f_A(W) - (W-\mu)f_A(W)\bigr| + 2\sigma^{-2},$$

$$d_{\mathrm{loc}}\bigl(\mathcal{L}(W),\mathrm{TP}\bigl(\mu,\sigma^2\bigr)\bigr) \le \sup_{a\in\mathbb{Z}^+}\bigl|\mathbb{E}\sigma^2\Delta f_a(W) - (W-\mu)f_a(W)\bigr| + 2\sigma^{-2}.$$

These inequalities form the basis of our approximations and leveraging them requires detailed understanding of the functions f_A. Though much is known about these functions due to their role in Stein's method for Poisson approximation (see, for example, [3]), our results require new, finer properties potentially of interest in other Poisson approximation settings; see Lemma 3.3.

3.2. Properties of the solutions of the Poisson Stein equation

We first review the known bounds on g_A.

Lemma 3.2 ([3]). *Let $A \subseteq \mathbb{Z}^+$, and let g_A be as in (3.1). Then*

$$\|g_A\| \le \frac{1}{\lambda^{1/2}} \quad \textit{and} \quad \|\Delta g_A\| \le \frac{1-e^{-\lambda}}{\lambda} \le \frac{1}{\lambda}.$$

If $A = \{a\}$ for some $a \in \mathbb{Z}^+$, then

$$\|g_a\| \le \frac{1}{\lambda}.$$

Note that f_A satisfies the same bounds as does g_A, but with $\lambda = \sigma^2 + \gamma$.

The bound on $\|g_a\|$ is smaller than the general bound on $\|g_A\|$ for $A \subset \mathbb{Z}^+$ by a factor of $\lambda^{-1/2}$. This suggests that the same might be true for a bound on $\|\Delta g_a\|$, but it is not the case: In fact, $\Delta g_a(a)$ is typically comparable to λ^{-1} and is not of order $\mathrm{O}(\lambda^{-3/2})$, as might have been hoped. In the remainder of this section, we establish non-uniform bounds on $|\Delta g_a(k)|$, showing that $|\Delta g_a(k)|$ is nonetheless 'typically' of order $\mathrm{O}(\lambda^{-3/2})$. This enables us to make the sharper local limit approximations of the paper.

Let $U_j := \{0, \dots, j-1\}$ and $\mathcal{P}_\lambda(\cdot) := \mathcal{P}(\lambda)\{\cdot\}$. Then the solution g_a of (3.1) with $A = \{a\}$ can be written as

$$g_a(k) = \lambda^{-k} e^{\lambda} (k-1)! \begin{cases} \mathcal{P}_\lambda(a)\mathcal{P}_\lambda(U_k^c), & k \ge a+1, \\ -\mathcal{P}_\lambda(a)\mathcal{P}_\lambda(U_k), & 1 \le k \le a. \end{cases} \tag{3.2}$$

We use this expression to prove the following bound.

Lemma 3.3. *Let g_a be as defined at (3.2). Then, for $k \ge 0$,*

$$\begin{aligned} \left|\Delta g_a(k)\right| \le{}& \frac{1}{\lambda^{3/2}\sqrt{2e}}\big(\mathrm{I}[k > a, k \ge \lambda] + \mathrm{I}[k < a, k < \lambda]\big) \\ &+ \left(\frac{\mathcal{P}_\lambda(a)}{a+1} + \frac{(\lambda - k)}{\lambda^2}\right)\mathrm{I}[a < k < \lambda] \\ &+ \left(\frac{\mathcal{P}_\lambda(a)}{\lambda} + \frac{(k-\lambda)}{\lambda^2}\right)\mathrm{I}[\lambda \le k < a] \\ &+ \frac{1}{\lambda}\mathrm{I}[k = a] \\ \le{}& \frac{1}{\lambda^{3/2}\sqrt{2e}} + \frac{|\lambda - k|}{\lambda^2} + \frac{1}{\lambda}\mathrm{I}[k = a]. \end{aligned}$$

Proof. The second bound follows from the first by noting that $\mathcal{P}_\lambda(a)/(a+1) = \lambda^{-1}\mathcal{P}_\lambda(a+1)$; here and below we use the bound $\sup_{k\ge 0}\mathcal{P}_\lambda(k)\sqrt{\lambda} \le (2e)^{-1/2}$, from [3], Proposition A.2.7.

The proof of the first bound consists of separate arguments in a number of cases, depending on the relative magnitudes of λ, k and a.

Case 1: Assume that $k \ge a+1$. Then

$$\begin{aligned}\Delta g_a(k) &= \lambda^{-k-1}e^\lambda (k-1)!\mathcal{P}_\lambda(a)\bigl(k\mathcal{P}_\lambda\bigl(U_{k+1}^c\bigr) - \lambda\mathcal{P}_\lambda\bigl(U_k^c\bigr)\bigr)\\ &= \lambda^{-k-1}e^\lambda (k-1)!\mathcal{P}_\lambda(a)\left(k\sum_{j=k+1}^\infty \frac{e^{-\lambda}\lambda^j}{j!} - \sum_{j=k+1}^\infty \frac{e^{-\lambda}\lambda^j}{(j-1)!}\right)\\ &= \lambda^{-k-1}e^\lambda (k-1)!\mathcal{P}_\lambda(a)\sum_{j=k+1}^\infty \frac{e^{-\lambda}\lambda^j}{j!}(k-j).\end{aligned}$$

We now use the fact that

$$I_k(\lambda) := \frac{1}{k!}\int_0^\lambda t^k e^{-t}\,dt = \sum_{j=k+1}^\infty \frac{e^{-\lambda}\lambda^j}{j!} = -\frac{e^{-\lambda}\lambda^k}{k!} + I_{k-1}(\lambda)$$

to give

$$\begin{aligned}\Delta g_a(k) &= \lambda^{-k-1}e^\lambda (k-1)!\mathcal{P}_\lambda(a)\bigl\{kI_k(\lambda) - \lambda I_{k-1}(\lambda)\bigr\}\\ &= -\frac{\mathcal{P}_\lambda(a)}{k}\left(1 + e^\lambda\lambda^{-k-1}(\lambda-k)\int_0^\lambda t^k e^{-t}\,dt\right)\\ &= -\frac{\mathcal{P}_\lambda(a)}{k}\left(1 + \frac{\lambda-k}{\lambda}\int_0^\lambda \left(\frac{t}{\lambda}\right)^k e^{\lambda-t}\,dt\right).\end{aligned} \tag{3.3}$$

Subcase 1.1: If $k \ge \lambda$, then

$$\begin{aligned}0 &\le \frac{k-\lambda}{\lambda}\int_0^\lambda \left(\frac{t}{\lambda}\right)^k e^{\lambda-t}\,dt = \frac{k-\lambda}{\lambda}\int_0^\lambda \left(1 + \frac{t-\lambda}{\lambda}\right)^k e^{\lambda-t}\,dt\\ &\le \frac{k-\lambda}{\lambda}\int_0^\lambda e^{(\lambda-t)(1-k/\lambda)}\,dt = 1 - e^{\lambda-k} \le 1.\end{aligned} \tag{3.4}$$

Therefore, in this case,

$$\begin{aligned}\left|1 - \frac{k-\lambda}{\lambda}\int_0^\lambda \left(\frac{t}{\lambda}\right)^k e^{\lambda-t}\,dt\right| &= 1 - \frac{k-\lambda}{\lambda}\int_0^\lambda \left(\frac{t}{\lambda}\right)^k e^{\lambda-t}\,dt\\ &\le 1 - \frac{k-\lambda}{\lambda}\int_0^\lambda \left(\frac{t}{\lambda}\right)^k dt = \frac{\lambda+1}{k+1},\end{aligned}$$

and so

$$\left|\Delta g_a(k)\right| \le \frac{\mathcal{P}_\lambda(a)(\lambda+1)}{k(k+1)} \le \frac{1}{\sqrt{2e}}\lambda^{-3/2}.$$

Subcase 1.2: If $a < k < \lambda$, then both summands in the expression for $\Delta g_a(k)$ at the end of (3.3) are positive, and so, because $I_k(\lambda) \le 1$, we have

$$\begin{aligned}\frac{\mathcal{P}_\lambda(a)}{k}\left(1+\frac{\lambda-k}{\lambda}\int_0^\lambda\left(\frac{t}{\lambda}\right)^k e^{\lambda-t}\,dt\right) &\le \frac{\mathcal{P}_\lambda(a)}{k}\left(1+\frac{(\lambda-k)k!}{e^{-\lambda}\lambda^{k+1}}\right)\\ &= \frac{\mathcal{P}_\lambda(a)}{k}+\frac{(\lambda-k)}{\lambda^2}\frac{\mathcal{P}_\lambda(a)}{\mathcal{P}_\lambda(k-1)} \qquad (3.5)\\ &\le \frac{\mathcal{P}_\lambda(a)}{a+1}+\frac{(\lambda-k)}{\lambda^2},\end{aligned}$$

where, in the last inequality, we have used the unimodality of the Poisson distribution.

Case 2: Assume that $k \le a-1$. Then, if $k \ge 1$, following arguments similar to those above, we have

$$\begin{aligned}\Delta g_a(k) &= \lambda^{-k-1}e^\lambda(k-1)!\mathcal{P}_\lambda(a)\big(\lambda\mathcal{P}_\lambda(U_k)-k\mathcal{P}_\lambda(U_{k+1})\big)\\ &= \lambda^{-k-1}e^\lambda(k-1)!\mathcal{P}_\lambda(a)\sum_{j=0}^{k-1}(j-k)\frac{e^{-\lambda}\lambda^j}{j!}\\ &= \frac{\mathcal{P}_\lambda(a)}{k}\left(-1+e^\lambda\lambda^{-k-1}(\lambda-k)\int_\lambda^\infty t^k e^{-t}dt\right)\\ &= \frac{\mathcal{P}_\lambda(a)}{k}\left(-1+\frac{\lambda-k}{\lambda}\int_\lambda^\infty\left(\frac{t}{\lambda}\right)^k e^{\lambda-t}\,dt\right).\end{aligned}$$

Subcase 2.1: If $k < \lambda$, we first take $k = 0$, where it is easy to see that

$$\left|\Delta g_a(0)\right| = -g_a(1) = \lambda^{-1}\mathcal{P}_\lambda(a) \le \frac{1}{\sqrt{2e}}\lambda^{-3/2}.$$

For $1 \le k < \lambda$, an argument similar to (3.4) shows that

$$\left|\Delta g_a(k)\right| = \frac{\mathcal{P}_\lambda(a)}{k}\left(1-\frac{\lambda-k}{\lambda}\int_\lambda^\infty\left(\frac{t}{\lambda}\right)^k e^{\lambda-t}dt\right),$$

and by bounding $t/\lambda > 1$, we easily find that

$$\left|\Delta g_a(k)\right| \le \frac{\mathcal{P}_\lambda(a)}{\lambda} \le \frac{1}{\sqrt{2e}}\lambda^{-3/2}.$$

Subcase 2.2: For $\lambda \le k < a$, following the same argument as in (3.5) yields that

$$\left|\Delta g_a(k)\right| \le \frac{\mathcal{P}_\lambda(a)}{\lambda} + \frac{(k-\lambda)}{\lambda^2} \le \frac{1}{\sqrt{2e}}\lambda^{-3/2} + \frac{(k-\lambda)}{\lambda^2}.$$

Case 3: If $k = a$, then we use the known bound $\|\Delta g_a\|_\infty < 1/\lambda$ from [3], Lemma 1.1.1. □

For translated Poisson approximation, the bound in Lemma 3.3 easily translates into the following result.

Lemma 3.4. *Let* $\mu, \sigma^2 > 0$, $s = \lfloor \mu - \sigma^2 \rfloor$, $\gamma = \mu - \sigma^2 - s$, *and set* $\lambda = \sigma^2 + \gamma$ *and* $f_a(k) = g_a(k-s)$ *for* g_a *the Poisson Stein solution defined at* (3.2). *Then*

$$\left|\Delta f_a(k)\right| \le \frac{1}{\sigma^3\sqrt{2e}} + \frac{|\mu - k|}{\sigma^4} + \frac{\mathrm{I}\{k = a + s\}}{\sigma^2}.$$

3.3. Completing the proof of Theorem 2.1

In order to exploit Proposition 3.1, we first need a manageable bound for the expectations $|\mathbb{E}\sigma^2\Delta f(W) - (W-\mu)f(W)|$ that appear there. This is given in the following lemma.

Lemma 3.5. *Let* (W, W', G, R) *be an approximate Stein coupling with* W *and* W' *integer valued*, $\mathbb{E}W = \mu$ *and* $\mathrm{Var}(W) = \sigma^2$. *Set* $D := W' - W$, *and let* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *be sigma-algebras such that* W *is* $\mathcal{F}_1$*-measurable and such that* (G, D) *is* $\mathcal{F}_2$*-measurable. Then*

$$\begin{aligned}
&\left|\mathbb{E}\sigma^2\Delta f(W) - (W-\mu)f(W)\right| \\
&\quad \le \left|\mathbb{E}\big[\big(\mathbb{E}[GD|\mathcal{F}_1] - \mathbb{E}GD\big)\Delta f(W)\big]\right| + \mathbb{E}\left|R(W-\mu)\right|\mathbb{E}\left|\Delta f(W)\right| + \mathbb{E}\left|Rf(W)\right| \qquad (3.6)\\
&\qquad + \mathbb{E}\left[\frac{|GD(D-1)|}{2}\min\big\{\|\Delta f\|S_1\big(\mathcal{L}(W|\mathcal{F}_2)\big), \|f\|S_2\big(\mathcal{L}(W|\mathcal{F}_2)\big)\big\}\right]. \qquad (3.7)
\end{aligned}$$

Proof. Since (W, W', G, R) is an approximate Stein coupling,

$$\mathbb{E}(W-\mu)f(W) = \mathbb{E}\big[G\big(f\big(W'\big) - f(W)\big)\big] - \mathbb{E}\big[Rf(W)\big],$$

and therefore

$$\begin{aligned}
&\left|\mathbb{E}\big[\sigma^2\Delta f(W) - (W-\mu)f(W)\big]\right| \\
&\quad \le \left|\mathbb{E}\big[\sigma^2\Delta f(W) - G\big(f\big(W'\big) - f(W)\big)\big]\right| + \mathbb{E}\left|Rf(W)\right|. \qquad (3.8)
\end{aligned}$$

With $D = W' - W$, add and subtract $GD\Delta f(W)$, and write $\sigma^2 = \mathbb{E}[GD] - \mathbb{E}\{R(W-\mu)\}$ using (1.8), giving

$$\begin{aligned}
\mathbb{E}\big[\sigma^2\Delta f(W) - G\big(f\big(W'\big) - f(W)\big)\big] = \mathbb{E}\big[&\big(\mathbb{E}[GD] - GD\big)\Delta f(W) - \mathbb{E}\big\{R(W-\mu)\big\}\Delta f(W) \\
&+ GD\Delta f(W) - G\big(f\big(W'\big) - f(W)\big)\big].
\end{aligned}$$

Hence (3.8) can be bounded by

$$\bigl|\mathbb{E}\bigl[\bigl(\mathbb{E}[GD|\mathcal{F}_1]-\mathbb{E}[GD]\bigr)\Delta f(W)\bigr]\bigr|+\mathbb{E}\bigl|R(W-\mu)\bigr|\mathbb{E}\bigl|\Delta f(W)\bigr|+\mathbb{E}\bigl|Rf(W)\bigr| \tag{3.9}$$

$$+\bigl|\mathbb{E}\bigl[GD\Delta f(W)-G\bigl(f\bigl(W'\bigr)-f(W)\bigr)\bigr]\bigr|. \tag{3.10}$$

It is easy to see that (3.9) is equal to (3.6). For (3.10), we observe that

$$\begin{aligned}
&D\Delta f(W)-\bigl(f\bigl(W'\bigr)-f(W)\bigr)\\
&\quad=\mathrm{I}[D>0]\sum_{i=0}^{D-1}\bigl(\Delta f(W)-\Delta f(W+i)\bigr)-\mathrm{I}[D<0]\sum_{i=1}^{-D}\bigl(\Delta f(W)-\Delta f(W-i)\bigr)\\
&\quad=-\mathrm{I}[D>1]\sum_{i=1}^{D-1}\sum_{j=0}^{i-1}\Delta^2 f(W+j)-\mathrm{I}[D<0]\sum_{i=1}^{-D}\sum_{j=0}^{i-1}\Delta^2 f(W-i+j).
\end{aligned}$$

Using this expression, conditioning on $\mathcal{F}_2$, and noting that for $k\in\mathbb{Z}$,

$$\bigl|\mathbb{E}\bigl[\Delta^2 f(W+k)|\mathcal{F}_2\bigr]\bigr|\le\min\bigl\{\|\Delta f\|S_1\bigl(\mathcal{L}(W|\mathcal{F}_2)\bigr),\|f\|S_2\bigl(\mathcal{L}(W|\mathcal{F}_2)\bigr)\bigr\},$$

we find that (3.7) upper bounds (3.10). □

Proof of Theorem 2.1. The total variation bound (2.1) is a consequence of Proposition 3.1 and Lemma 3.5, together with the bounds from Lemma 3.2; cf. [31], Theorem 3.1, and [20], Theorem 1.3. To prove (2.2), we can argue similarly, but taking $f=f_a$ and using Lemma 3.4 to bound $\Delta f_a(\cdot)$. This yields (2.2): the terms in (2.2) (except for the last one) bound (3.6), and the last term in (2.2) bounds (3.7) and the extra term $2\sigma^{-2}$ in Proposition 3.1. □

4. Proofs of applications

In this section, we prove the application results given in Section 2.

4.1. Hoeffding combinatorial local central limit theorem

Recall the definitions of a, W, μ and σ^2 from Section 2 and also Assumptions A1 and A2.

The first step is to find a Stein coupling for W. We choose among those discussed in [13], Section 4.1. The most direct procedure is to let I, J be i.i.d. uniform on $\{1,\dots,n\}$, and to define $W':=W-a_{I\rho_I}-a_{J\rho_J}+a_{I\rho_J}+a_{J\rho_I}$; then (W,W') is an exchangeable pair satisfying $\mathbb{E}[W'-W|W]=-\frac{2}{n-1}(W-\mu)$, giving an exact Stein coupling with $G:=(n-1)(W'-W)/4$. However, we use another exact Stein coupling, that yields a simpler form for T in (2.4); we take

$$W':=W-a_{I\rho_I}-a_{J\rho_J},\qquad I\ne J;\qquad W':=W-a_{I\rho_I},\qquad I=J;\qquad G:=n(a_{I\rho_J}-a_{I\rho_I}).$$

Now, $\mathbb{E}[GD|\rho] = 2n^{-1}\mu(W-\mu) + \sum_{l=1}^{4} T_l + n^{-1}\mu^2$, where

$$T_1 := \sum_i a_{i\rho_i}^2, \qquad T_2 := -\frac{1}{n}\sum_i a_{i\rho_i} a_{i+}, \qquad T_3 := -\frac{1}{n}\sum a_{+\rho_j} a_{j\rho_j}, \qquad T_4 := \frac{1}{n}(W-\mu)^2,$$

and so

$$\bigl|\mathbb{E}[GD|\rho] - \mathbb{E}[GD]\bigr| \le 2A_1|W-\mu| + T, \qquad \text{with } T := \sum_{l=1}^{4} |T_l - \mathbb{E}T_l|,$$

satisfying condition (2.4) with $k=1$ and $\kappa = 2A_1$. We now consider the remaining conditions to be satisfied in Corollary 2.8(i), (ii). Since $R=0$, for (i), we need to bound $\mathbb{E}T$ and Υ. Note that since $D := W' - W$ satisfies $|D| \le 2A_1$, and also that $\sigma^{-2}|G| \le 2/(\alpha_0^2 A_1)$, we have $\Upsilon \le C\sigma^2 \mathbb{E}[S_2(\mathcal{L}(W)|\mathcal{F}_2)]$, where we define $\mathcal{F}_2$ be the sigma-algebra generated by (I, J, ρ_I, ρ_J).

The smoothing term. We begin with the smoothing coefficient $\mathbb{E}[S_2(\mathcal{L}(W)|\mathcal{F}_2)]$; recall Assumption A2.

Lemma 4.1. *Under Assumptions* A1 *and* A2, *we have*

$$\mathbb{E}\bigl[S_2\bigl(\mathcal{L}(W)|\mathcal{F}_2\bigr)\bigr] = \mathrm{O}\bigl(n^{-1}\bigr).$$

Proof. Condition on I, J, ρ_I, ρ_J, let ρ° be a uniformly chosen permutation given $\rho_k^\circ = \rho_k$ for $k = I, J$, and define $W^\circ = \sum_{i=1}^n a_{i\rho_i^\circ}$ so that $\mathcal{L}(W^\circ) = \mathcal{L}(W|\mathcal{F}_2)$. Now, for each $1 \le l \le n_1$ such that neither of i_{l1}, i_{l2} are equal to I or J, independently and with probability $1/2$, multiply ρ° by the transposition $(\rho_{i_{l1}}^\circ, \rho_{i_{l2}}^\circ)$, so that, if the multiplication takes place, then $i_{l2} \mapsto \rho_{i_{l1}}^\circ$ and $i_{l1} \mapsto \rho_{i_{l2}}^\circ$. This process forms a new permutation $\tilde{\rho}$, which is still uniformly distributed given $\tilde{\rho}_k = \rho_k$ for $k = I, J$, so that $\widetilde{W} := \sum_{i=1}^n a_{i\tilde{\rho}_i}$ has the same distribution as W°. Moreover, writing

$$C_l\bigl(\rho^\circ\bigr) := (a_{i_{l1},\rho_{i_{l2}}^\circ} + a_{i_{l2},\rho_{i_{l1}}^\circ} - a_{i_{l1},\rho_{i_{l1}}^\circ} - a_{i_{l2},\rho_{i_{l2}}^\circ}),$$

we have for $E_{IJ} := \{l : 1 \le l \le n_1, \{i_{l1}, i_{l2}\} \cap \{I, J\} = \varnothing\}$,

$$\widetilde{W} = W^\circ + \sum_{l \in E_{IJ}} B_l C_l\bigl(\rho^\circ\bigr),$$

where $B_1, \dots, B_{n_1}$ are i.i.d. Bernoulli Be$(1/2)$ random variables, independent of ρ. Defining

$$N\bigl(\rho^\circ\bigr) := \bigl|\bigl\{l : \{i_{l1}, i_{l2}\} \cap \{I, J\} = \varnothing, \bigl|C_l\bigl(\rho^\circ\bigr)\bigr| = 1\bigr\}\bigr|,$$

it thus follows immediately that, on the event $\{N(\rho^\circ) \ge k_n\}$, we have

$$S_2\bigl(\mathcal{L}\bigl(\widetilde{W}|\rho^\circ\bigr)\bigr) \le S_2\bigl(\mathrm{Bi}(k_n, 1/2)\bigr) \le 10k_n^{-1},$$

where the last inequality is [34], Proposition 3.8.

Taking expectations, this in turn implies that

$$S_2\big(\mathcal{L}(W|\mathcal{F}_2)\big) = S_2\big(\mathcal{L}\big(W^\circ\big)\big) = S_2\big(\mathcal{L}(\widetilde{W})\big) \le 8k_n^{-1} + 4\mathbb{P}\big(N\big(\rho^\circ\big) < k_n\big).$$

Defining $k_n := \lfloor \frac{1}{2}\mathbb{E}N(\rho^\circ)\rfloor$, we show that, for suitable $\beta_1 > 0$, $\beta_2 < \infty$, we have $k_n \ge \beta_1 n$ and $\mathbb{P}(N(\rho^\circ) < k_n) \le \beta_2 n^{-1}$ for large n, thus completing the proof of the lemma.

Note that ρ° is a uniformly chosen map from $\{1,\dots,n\}\setminus\{I,J\}$ to $\{1,\dots,n\}\setminus\{\rho_I,\rho_J\}$. Thus, from our assumption on the matrix a,

$$\begin{aligned}
\mathbb{E}N\big(\rho^\circ\big) &\ge -2 + \sum_{l\in E_{IJ}} \mathbb{P}\big(\big|C_i\big(\rho^\circ\big)\big| = 1\big) \\
&= -2 + \sum_{l\in E_{IJ}} \mathbb{P}\big(\big\{\rho^\circ_{i_{l1}}, \rho^\circ_{i_{l2}}\big\} \in \mathcal{J}(i_{l1}, i_{l2})\big) \\
&\ge -2 + \frac{(n_1-2)(n_2-2n)}{n(n-1)} \ge n\left(\left(\alpha_1 - \frac{2}{n}\right)\left(\alpha_2 - \frac{2}{n}\right) - \frac{2}{n}\right),
\end{aligned}$$

and so $k_n \ge \beta_1 n$ for n large, with $\beta_1 := \alpha_1\alpha_2/4$. Then, by Chebyshev's inequality, we have the upper bound

$$\mathbb{P}\big(N\big(\rho^\circ\big) \le k_n\big) \le \frac{\operatorname{Var}(N(\rho^\circ))}{(\mathbb{E}N(\rho^\circ) - k_n)^2} = \frac{4\operatorname{Var}(N(\rho^\circ))}{(\mathbb{E}N(\rho^\circ))^2}. \tag{4.1}$$

It is now enough to show that $\operatorname{Var}(N(\rho^\circ)) \le Cn$, for some $C < \infty$.

To bound $\operatorname{Var}(N(\rho^\circ))$, we need to compute the covariance of X_l and X_k for $l \ne k$, both in E_{IJ}, where

$$X_r := \mathrm{I}\big(\big\{\rho^\circ_{i_{r1}}, \rho^\circ_{i_{r2}}\big\} \in \mathcal{J}(i_{r1}, i_{r2})\big).$$

Now, given $X_l = 1$ and $(\rho^\circ_{i_{l1}}, \rho^\circ_{i_{l2}}) = (j_1, j_2)$, the conditional probability that $X_k = 1$ can be no larger that $\mathbb{P}(X_k = 1)(n-2)(n-3)/(n-4)(n-5)$, since pairs that are excluded by having (j_1, j_2) as images of i_{l1} and i_{l2} under ρ would only reduce the conditional probability, and the probability of an accessible pair being attained is increased by the factor $(n-2)(n-3)/(n-4)(n-5)$. Hence,

$$\begin{aligned}
\operatorname{Cov}(X_l, X_k) &\le \mathbb{P}(X_l = 1)\mathbb{P}(X_k = 1)\left\{\frac{(n-2)(n-3)}{(n-4)(n-5)} - 1\right\} \\
&\le \mathbb{P}(X_l = 1)\mathbb{P}(X_k = 1)\frac{2(2n-7)}{(n-4)(n-5)},
\end{aligned}$$

so that, for $n \ge 28$,

$$\begin{aligned}
\operatorname{Var}\big(N\big(\rho^\circ\big)\big) &= \sum_{l\in E_{IJ}} \operatorname{Var}(X_l) + \sum_{l\ne k;\, l,k\in E_{IJ}} \operatorname{Cov}(X_l, X_k) \\
&\le \mathbb{E}N\big(\rho^\circ\big) + 5n^{-1}\big\{\mathbb{E}N\big(\rho^\circ\big)\big\}^2 = \mathrm{O}(n).
\end{aligned}$$

This proves the lemma. □

As a consequence of Lemma 4.1 and the remarks preceding it, we have $\Upsilon = \mathrm{O}(1)$, and we now show $\mathbb{E}T = \mathrm{O}(\sigma)$, after which, Corollary 2.8(i) implies

$$d_{\mathrm{TV}}\big(\mathcal{L}(W), \mathrm{TP}\big(\mathbb{E}W, \mathrm{Var}(W)\big)\big) = \mathrm{O}(1/\sigma). \tag{4.2}$$

Observing first that

$$\mathbb{E}\big[T\mathrm{I}[W=k]\big] = \sum_{l=1}^{4} \mathbb{E}\big[|T_l - \mathbb{E}T_l|\mathrm{I}[W=k]\big],$$

we apply Lemma 2.6 to the first three terms; for the fourth, we immediately have

$$\mathbb{E}\big[|T_4 - \mathbb{E}T_4|\mathrm{I}(W=k)\big] \le 2\mathbb{E}T_4 = n^{-1}\sigma^2 = O(1),$$

so that this element of T gives a contribution to the error bound of order $\mathrm{O}(\sigma^{-2})$.

The remaining elements each have the form $|\sum_i \tilde{a}_{i\rho_i} - \mathbb{E}\sum_i \tilde{a}_{i\rho_i}|$, for appropriate choices of $\tilde{a}_{ij} \le A_1^2$ (take $\tilde{a}_{ij}$ to be a_{ij}^2, $a_{ij}a_{i+}/n$ and $a_{ij}a_{+j}/n$, respectively). It thus follows from (2.13) that

$$\mathbb{E}|T_l - \mathbb{E}T_l| \le \sqrt{\mathrm{Var}\, T_l} = \mathrm{O}(\sqrt{n}), \qquad 1 \le l \le 3,$$

and hence that $\mathbb{E}T = \mathrm{O}(\sigma)$, as desired.

For the local approximation, Lemma 4.1 and (4.2) imply that Condition (1) of of Corollary 2.8(ii) is satisfied with $\alpha = 1$. For Condition (2), under Assumption A1, [8], Proposition 1.2, implies that, for any $j \ge 0$,

$$\sigma^{-2j}\mathbb{E}(W-\mu)^{2j} \le (2j-1)^{2j} A_1^{2j} \sigma^{-2j} n^j = \mathrm{O}(1).$$

Finally, we (essentially) use Condition (3b) for the T term and treat T_1, T_2 and T_3 using concentration bounds. Under Assumption A1, [8], Proposition 1.1, and [23], Theorem 3.1, imply, for $l = 1, 2, 3$, that

$$\mathbb{P}\big((A_1\sigma)^{-1}|T_l - \mathbb{E}T_l| \ge t\big) \le 2\exp\left\{-\frac{t^2}{2(v_l/(A_1\sigma)^2) + 16(A_1/\sigma)t}\right\},$$

where $v_l = \mathrm{Var}(T_l)$, and $v_l/(A_1\sigma)^2 \le C_l$ for some suitable C_l. We can then apply Lemma 2.6, taking

$$\varepsilon_l(t) = t\overline{F}_l(t) + \int_t^\infty \overline{F}_l(v)\,dv,$$

where

$$\overline{F}_l(t) := 2\exp\left\{-\frac{t^2}{2C_l + (16/\alpha_0)tn^{-1/2}}\right\},$$

and the choice $t = C_l'\sqrt{\log\sigma}$, for C_l' suitably large but fixed, gives

$$\mathbb{E}\big[|T_l - \mathbb{E}T_l|\mathrm{I}[W=k]\big] \le \sigma\varepsilon_l(t) + t = \mathrm{O}(\sqrt{\log\sigma}).$$

Hence, from Corollary 2.8, under Assumptions A1 and A2, we have

$$d_{\mathrm{loc}}\big(\mathcal{L}(W), \mathrm{TP}\big(\mathbb{E}W, \mathrm{Var}(W)\big)\big) = \mathrm{O}\left(\frac{\sqrt{\log(\sigma)}}{\sigma^2}\right).$$

4.2. Number of isolated vertices in an Erdős–Rényi random graph

Let $\mathcal{G} := \mathcal{G}(n, p)$ be an Erdős–Rényi random graph on n vertices $v_1, \dots, v_n$, and let W be the number of isolated vertices in $\mathcal{G}$. Let W^s have the size-biased distribution of W. Then $(W, W', G) = (W, W^s, \mathbb{E}W)$ is a Stein coupling. To couple (W, W^s), construct W^s from $\mathcal{G}$ by choosing a vertex at random and erasing all edges (if any) connected to the vertex. Recall from the introduction that we consider the regime $p \asymp \lambda/n$ for some $\lambda > 0$, in which case $\mu \sim ne^{-\lambda}$ and $\sigma^2 \sim ne^{-\lambda}\{1 + (\lambda - 1)e^{-\lambda}\}$ are of strict order n.

Let E_i be the event that vertex v_i is not isolated in $\mathcal{G}$, let $W_1(v)$ be the number of degree-1 vertices connected to vertex v of $\mathcal{G}$, and let W_1 be the number of degree-1 vertices in $\mathcal{G}$. To check (2.4), we observe that

$$\begin{aligned}\big|\mathbb{E}[GD|\mathcal{G}] - \sigma^2\big| &= \left|\frac{\mu}{n}\sum_{i=1}^{n}\big(W_1(v_i) + \mathrm{I}\{E_i\}\big) - \sigma^2\right| \\ &= \left|\frac{\mu}{n}\big(W_1 + (n - W)\big) - \sigma^2\right|.\end{aligned}$$

Since $\mathbb{E}[GD] = \sigma^2$, which is (1.8) for a Stein coupling (thus with $R = 0$), it follows that

$$\begin{aligned}\big|\mathbb{E}[GD|\mathcal{G}] - \sigma^2\big| &= \left|\frac{\mu}{n}\big((W_1 - \mathbb{E}W_1) - (W - \mathbb{E}W)\big)\right| \\ &\le (1-p)^{n-1}\big\{|W - \mu| + |W_1 - \mathbb{E}W_1|\big\} \\ &\le |W - \mu| + |W_1 - \mathbb{E}W_1|,\end{aligned}$$

which is (2.4) with $k = 1$, $\kappa = 1$ and $T = |W_1 - \mathbb{E}W_1|$.

To apply Corollary 2.8, we need to show that $\sigma^{-j}\mathbb{E}|W - \mu|^j = \mathrm{O}(1)$ for suitable values of j. To do so, and also to show that the distribution of T is concentrated, we take $d = 0$ and $d = 1$ in the following theorem of Bartroff, Goldstein and Işlak [6] (see also [1]).

Theorem 4.2. *For any integer $d \ge 0$, let W_d be the number of degree d vertices in an Erdős–Rényi random graph $\mathcal{G}$ with parameters n and p. Then, for any $t > 0$,*

$$\mathbb{P}\big(|W_d - \mathbb{E}W_d| > t\big) \le 2\exp\left\{-\frac{t^2}{4(n - \mathbb{E}W_d) + (4/3)t}\right\}.$$

So, for any value of d, we have

$$\mathbb{P}\big(\sigma^{-1}|W_d - \mathbb{E}W_d| \geq t\big) \leq 2\exp\left\{-\frac{t^2}{4\frac{n}{\sigma^2} + \frac{4}{3\sigma}t}\right\}$$
$$\leq \eta_n(t) := 2\exp\left\{-\frac{t^2}{4\gamma + 4t\sqrt{\gamma/n}/3}\right\},$$

where γ is an upper bound for n/σ^2. It follows easily, taking $d = 0$, that $\sigma^{-k}\mathbb{E}|W - \mu|^k = \mathrm{O}(1)$ for all $k \geq 1$, and then, taking $d = 1$ and

$$\varepsilon_n(t) := t\eta_n(t) + \int_t^\infty \eta_n(v)\,dv,$$

that $\int_1^\infty \varepsilon_n(t)\,dt \leq \int_1^\infty \varepsilon_1(t)\,dt < \infty$ for all $n \geq 1$, so that $\mathbb{E}T^2 = \mathrm{O}(\sigma^2)$. Since also $R = 0$ almost surely, and since [34], Lemma 4.7, shows that

$$S_2\big(\mathcal{L}(W)\big) = \mathrm{O}\big(\sigma^{-2}\big),$$

then once we show $\Upsilon = \mathrm{O}(1)$, all the hypotheses and conditions of Corollary 2.8(i), (ii) except for (3) of (ii) are satisfied, with $\alpha = 1$. But a variation of Condition (3b) is satisfied: for $t = t_n = c\sqrt{\log\sigma}$, it is easy to check that $\varepsilon_n(t_n) = \mathrm{O}(n^{-1/2})$ if $c > 0$ is chosen fixed but large enough, so that, from Lemma 2.6, $\mathbb{E}[T\mathrm{I}[W = k]] = \mathrm{O}(\sqrt{\log\sigma})$, and thus the contribution from T is at most of order $\mathrm{O}(\sigma^{-2}\sqrt{\log\sigma})$.

All that is left is to to show that $\Upsilon = \mathrm{O}(1)$, for which we follow [20]. Let I, uniformly distributed on $\{1, \ldots, n\}$, be the index of the vertex of $\mathcal{G}$ chosen to be isolated in constructing W^s, and for $k = 0, 1, 2$, let $\mathcal{N}_k^{(i)}$ be the set of vertices at distance k from vertex v_i in $\mathcal{G}$. Then let $\mathcal{F}_2$ be the sigma algebra generated by $(I, \mathcal{N}_1^{(I)}, \mathcal{N}_2^{(I)})$ and the presence or absence of all edges that have one or more vertices in $\{I\} \cup \mathcal{N}_1^{(I)}$. Clearly $W^s - W$ is $\mathcal{F}_2$-measurable. To bound $\Upsilon := \mathbb{E}[|GD(D-1)|S_2(W|\mathcal{F}_2)]$, consider the expectation on the event $\{|\mathcal{N}_1^{(I)}| > \sqrt{n}\}$ and on its complement.

First, note that $|D| = |W^s - W| \leq 1 + |\mathcal{N}_1^{(I)}|$, so that

$$\mathbb{E}\big[|GD(D-1)|S_2(W|\mathcal{F}_2)\mathrm{I}\big\{|\mathcal{N}_1^{(I)}| > \sqrt{n}\big\}\big] \leq Cn\mathbb{E}\big[\big(1 + |\mathcal{N}_1^{(I)}|^2\big)\mathrm{I}\big\{|\mathcal{N}_1^{(I)}| > \sqrt{n}\big\}\big].$$

Since $|\mathcal{N}_1^{(I)}| \sim \mathrm{Bi}(n-1, p)$, and $p \sim \lambda/n$,

$$\mathbb{E}\big[|\mathcal{N}_1^{(I)}|^k\big] \leq C_k \qquad \text{for all } n \geq 1,$$

for suitable constants C_k, so that

$$\mathbb{E}\big[\big(1 + |\mathcal{N}_1^{(I)}|^2\big)\mathrm{I}\big\{|\mathcal{N}_1^{(I)}| > \sqrt{n}\big\}\big] = \mathrm{O}\big(n^{-k/2}\big)$$

for all integers $k \geq 1$.

For the complementary event, we show that, for some universal constant C,

$$S_2(W|\mathcal{F}_2)\mathrm{I}\{|\mathcal{N}_1^{(I)}| \le \sqrt{n}\} \le C\sigma^{-2}, \qquad \text{a.s.} \tag{4.3}$$

If this is the case, then

$$\mathbb{E}\big[|GD(D-1)|S_2(W|\mathcal{F}_2)\mathrm{I}\{|\mathcal{N}_1^{(I)}| \le \sqrt{n}\}\big] \le Cn\sigma^{-2}\mathbb{E}\big[\big(1+|\mathcal{N}_1^{(I)}|^2\big)\big] = \mathrm{O}(1),$$

as desired. For (4.3), the basic idea is that there still remain almost $\binom{n}{2}$ edges to be independently assigned, and the methods leading to [34], Lemma 4.7(i), can be applied to give the required order.

From now on, we have $|\mathcal{N}_1^{(I)}| \le \sqrt{n}$. Given $\mathcal{F}_2$, define a new random graph $\widetilde{\mathcal{G}}$ on n vertices labeled $\{v_1, \dots, v_n\}$ such that all edges with an endpoint in $V(I) := \{v_i : i \in \{I\} \cup \mathcal{N}_1^{(I)}\}$ are determined by $\mathcal{F}_2$, and the remaining edges, those in $E(I) := \{\{i, j\} : i, j \notin V(I)\}$, are assigned using i.i.d. Be(p) variables; we let $\widetilde{\mathcal{G}}(I)$ denote the graph $\widetilde{\mathcal{G}}$ restricted to $E(I)$. Note that the number of edges in $E(I)$ is

$$\binom{n - |\mathcal{N}_1^{(I)}| - 1}{2} \sim \frac{1}{2}n^2,$$

because $|\mathcal{N}_1^{(I)}| \le \sqrt{n}$. Let $\widetilde{\mathcal{G}}'$ be the graph obtained by choosing at random one of the edges of $E(I)$ and resampling it, and let $\widetilde{\mathcal{G}}''$ be the graph obtained from the same operation applied to $\widetilde{\mathcal{G}}'$. Let $\widetilde{W}, \widetilde{W}', \widetilde{W}''$ be the number of isolated vertices in $\widetilde{\mathcal{G}}, \widetilde{\mathcal{G}}', \widetilde{\mathcal{G}}''$. Then $\mathcal{L}(\widetilde{W}) = \mathcal{L}(W|\mathcal{F}_2)$, and $(\widetilde{W}, \widetilde{W}', \widetilde{W}'')$ are three successive states of a reversible Markov chain. Thus [34], Theorem 3.7, implies that

$$\begin{aligned}
&S_2\big(\mathcal{L}(W|\mathcal{F}_2)\big) \\
&\quad\le \frac{1}{\mathbb{P}(\widetilde{W}' = \widetilde{W}+1)^2}\big[2\operatorname{Var}\big(\mathbb{P}\big(\widetilde{W}' = \widetilde{W}+1|\tilde{\mathcal{G}}\big)\big) + 2\operatorname{Var}\big(\mathbb{P}\big(\widetilde{W}' = \widetilde{W}-1|\tilde{\mathcal{G}}\big)\big) \\
&\qquad + \mathbb{E}\big|\mathbb{P}\big(\widetilde{W}'' = \widetilde{W}'+1, \widetilde{W}' = \widetilde{W}+1|\tilde{\mathcal{G}}\big) - \mathbb{P}\big(\widetilde{W}' = \widetilde{W}+1|\tilde{\mathcal{G}}\big)^2\big| \\
&\qquad + \mathbb{E}\big|\mathbb{P}\big(\widetilde{W}'' = \widetilde{W}'-1, \widetilde{W}' = \widetilde{W}-1|\tilde{\mathcal{G}}\big) - \mathbb{P}\big(\widetilde{W}' = \widetilde{W}-1|\tilde{\mathcal{G}}\big)^2\big|\big].
\end{aligned}$$

Bounds on the first two terms are given by [20], Inequalities (2.23)–(2.25), which yield

$$\mathbb{P}\big(\widetilde{W}' = \widetilde{W}+1\big) \ge Cn^{-1} \quad \text{and} \quad \operatorname{Var}\big(\mathbb{P}\big(\widetilde{W}' = \widetilde{W} \pm 1|\tilde{\mathcal{G}}\big)\big) \le Cn^{-3}.$$

For the last two terms of the bound, let $\mathcal{V}_1^{(I)}$ be the set of vertices having degree one in both of $\widetilde{\mathcal{G}}$ and $\widetilde{\mathcal{G}}(I)$, and let $\widehat{\mathcal{V}}_1^{(I)}$ be the subset of these vertices that are connected to a vertex having degree two in both of $\widetilde{\mathcal{G}}$ and $\widetilde{\mathcal{G}}(I)$; write $\widetilde{W}_1^{(I)} := |\mathcal{V}_1^{(I)}|$ and $\widehat{W}_1^{(I)} := |\widehat{\mathcal{V}}_1^{(I)}|$. Let $\mathcal{E}_2^{(I)}$ be the set of edges that are isolated in both $\widetilde{\mathcal{G}}$ and $\widetilde{\mathcal{G}}(I)$, and let $\mathcal{E}_3^{(I)}$ be the set of pairs of connected edges that are isolated in both $\widetilde{\mathcal{G}}$ and $\widetilde{\mathcal{G}}(I)$; denote their numbers by $E_2^{(I)}$ and $E_3^{(I)}$ respectively. Note that no vertices of $N_1^{(I)}$ are isolated, but that v_I may be isolated (and then $N_1^{(I)}$ is empty); note also that the endpoints of elements of $\mathcal{E}_3^{(I)}$ belong to $\widehat{\mathcal{V}}_1^{(I)}$.

Now the only way to increase the number of isolated vertices in going from $\widetilde{\mathcal{G}}$ to $\widetilde{\mathcal{G}}'$ is to choose a non-isolated edge connected to a degree one vertex, and then remove it; however, the number of isolated vertices increases by 2 if the edge removed belongs to $\mathcal{E}_2^{(I)}$. Hence, writing $\hat{n}(I) := n - |\mathcal{N}_1^{(I)}| - 1$, we have

$$\mathbb{P}\big(\widetilde{W}' = \widetilde{W} + 1 | \widetilde{\mathcal{G}}\big) = \frac{(\widetilde{W}_1^{(I)} - 2\widetilde{E}_2^{(I)})}{\binom{\hat{n}(I)}{2}}(1-p).$$

Considering the different ways of increasing the number of isolated vertices by exactly one in consecutive steps is more complicated; isolating a vertex in $\widehat{\mathcal{V}}_1^{(I)}$ leaves the number of vertices of degree 1 unchanged, so that $(\widetilde{W}_1^{(I)})' = \widetilde{W}_1^{(I)}$, but if the vertex belonged to an element of $\mathcal{E}_3^{(I)}$, then $(\widetilde{E}_2^{(I)})' = \widetilde{E}_2^{(I)} + 1$. Hence

$$\begin{aligned}\mathbb{P}\big(\widetilde{W}'' = \widetilde{W}' + 1, \widetilde{W}' = \widetilde{W} + 1 | \widetilde{\mathcal{G}}\big) &= \frac{(\widetilde{W}_1^{(I)} - 2\tilde{E}_2^{(I)} - \tilde{V}_1^{(I)})(\widetilde{W}_1^{(I)} - 1 - 2\tilde{E}_2^{(I)})}{\binom{\hat{n}(I)}{2}^2}(1-p)^2 \\ &\quad + \frac{(\tilde{V}_1^{(I)} - 2\tilde{E}_3^{(I)})(\widetilde{W}_1^{(I)} - 2\tilde{E}_2^{(I)})}{\binom{\hat{n}(I)}{2}^2}(1-p)^2 \\ &\quad + \frac{2\tilde{E}_3^{(I)}(\widetilde{W}_1^{(I)} - 2\tilde{E}_2^{(I)} - 2)}{\binom{\hat{n}(I)}{2}^2}(1-p)^2,\end{aligned}$$

so that

$$\begin{aligned}&\mathbb{E}\big|\mathbb{P}\big(\widetilde{W}'' = \widetilde{W}' + 1, \widetilde{W}' = \widetilde{W} + 1 | \widetilde{\mathcal{G}}\big) - \mathbb{P}\big(\widetilde{W}' = \widetilde{W} + 1 | \widetilde{\mathcal{G}}\big)^2\big| \\ &\quad \leq \frac{\mathbb{E}|\widetilde{W}_1^{(I)} - 2\widetilde{E}_2^{(I)}| + \mathbb{E}\widehat{W}_1^{(I)} + 4\mathbb{E}\widetilde{E}_3^{(I)}}{\binom{\hat{n}(I)}{2}^2}(1-p)^2 = \mathrm{O}\big(n^{-3}\big),\end{aligned}$$

since $|\widetilde{W}_1^{(I)} - 2\widetilde{E}_2^{(I)}| \leq \widetilde{W}_1^{(I)} \leq n$, $\widehat{W}_1^{(I)} \leq n$, $\widetilde{E}_3^{(I)} \leq n$, and $|\mathcal{N}_1^{(I)}| \leq \sqrt{n}$. Similarly, but more easily, we have

$$\mathbb{P}\big(\widetilde{W}' = \widetilde{W} - 1 | \widetilde{\mathcal{G}}\big) = \frac{(\widetilde{W} - \mathrm{I}[\deg(v_I) = 0])(\hat{n}(I) - \widetilde{W} + \mathrm{I}[\deg(v_I) = 0])}{\binom{\hat{n}(I)}{2}}p$$

and

$$\mathbb{P}\big(\widetilde{W}'' = \widetilde{W}' - 1, \widetilde{W}' = \widetilde{W} - 1 | \widetilde{\mathcal{G}}\big) = \frac{4\binom{\widetilde{W} - \mathrm{I}[\deg(v_I)=0]}{2}\binom{\hat{n}(I) - \widetilde{W} + \mathrm{I}[\deg(v_I)=0]+1}{2}}{\binom{\hat{n}(I)}{2}^2}p^2,$$

so that

$$\mathbb{E}\big|\mathbb{P}\big(\widetilde{W}'' = \widetilde{W}' - 1, \widetilde{W}' = \widetilde{W} - 1|\widetilde{\mathcal{G}}\big) - \mathbb{P}\big(\widetilde{W}' = \widetilde{W} - 1|\widetilde{\mathcal{G}}\big)^2\big|$$
$$= \frac{(\widetilde{W} - \mathrm{I}[\deg(v_I) = 0])(\hat{n}(I) - \widetilde{W} + \mathrm{I}[\deg(v_I) = 0])(\hat{n}(I) - 2\widetilde{W} + 2\mathrm{I}[\deg(v_I) = 0] + 1)p^2}{\binom{\hat{n}(I)}{2}^2},$$

which is again $\mathrm{O}(n^{-3})$, since $0 \le \widetilde{W} \le n$. Therefore, $S_2(\mathcal{L}(W|\mathcal{F}_2)) = \mathrm{O}(n^{-1})$ almost surely, as desired.

4.3. Curie–Weiss

Recall from Section 2 the definition of the Curie–Weiss distribution, the magnetization W, and associated discussion. Assume that either $h > 0$ and $\beta > 0$, or that $h = 0$ and $0 < \beta < 1$. Define the exchangeable pair (W, W') as follows. Let I be uniform on $\{1, \dots, n\}$. Given $I = i$ and $S = s$, let

$$\mathbb{P}\big(S_i' = x\big) = \mathbb{P}\big(S_i = x|(S_j)_{j\neq i} = (s_j)_{j\neq i}\big)$$

for $x = \pm 1$. Defining $W' := W - S_I + S_I'$, we note that (W, W') are two consecutive states of a stationary Gibbs sampler, and so form an exchangeable pair. Note that W actually sits on a lattice of span 2 (even or odd numbers, depending on n), so that our eventual conclusion concerns $\widetilde{W} := (W + \frac{1}{2}\{1 - (-1)^n\})/2$.

We next want to establish an approximate linear regression, so as to determine an approximate Stein coupling. From [8], page 315 (see also [35], (7.10)), for $\mu_n := \mathbb{E}W$, we have

$$\begin{aligned}
\mathbb{E}\big[W' - W|S\big] &= -\frac{1}{n}W + \frac{1}{n}\sum_{i=1}^n \tanh\left(\frac{\beta}{n}(W - S_i) + h\right) \\
&= -\frac{1}{n}(W - \mu_n) + \frac{1}{n}\sum_{i=1}^n\left(\tanh\left(\frac{\beta}{n}(W - S_i) + h\right) - \tanh\left(\frac{\beta}{n}W + h\right)\right) \qquad (4.4)\\
&\quad + \tanh\left(\frac{\beta}{n}W + h\right) - \tanh(\beta m_h + h) + (m_h - \mu_n/n).
\end{aligned}$$

Now since, $0 \le \frac{d}{dx}\tanh(x) = 1 - \tanh^2(x) \le 1$ and $|\frac{d^2}{dx^2}\tanh(x)| \le 1$, by Taylor expansion we have

$$\big|\tanh(\beta w + h) - \tanh(\beta m + h) - \beta(w - m)\big(1 - \tanh^2(\beta m + h)\big)\big| \le C|w - m|^2;$$
$$\big|\tanh\big(\beta(w - s) + h\big) - \tanh(\beta w + h)\big| \le \beta s,$$

from which it follows that

$$\left|\tanh\left(\frac{\beta}{n}W + h\right) - \tanh(\beta m_h + h) - n^{-1}\beta(W - nm_h)\big(1 - m_h^2\big)\right| \le C\left(\frac{W}{n} - m_h\right)^2 \qquad (4.5)$$

and

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(\tanh\left(\frac{\beta}{n}(W-S_i)+h\right)-\tanh\left(\frac{\beta}{n}W+h\right)\right)\right|\le\frac{\beta}{n}. \tag{4.6}$$

This gives an approximate linear regression $\mathbb{E}[W'-W|W]=-a(W-\mu_n)+aR$, with

$$\begin{aligned} &a:=n^{-1}\big(1-\beta\big(1-m_h^2\big)\big);\\ &|R|\le R':=\frac{\beta}{1-\beta(1-m_h^2)}+|\mu_n-nm_h|+\frac{Cn|W/n-m_h|^2}{1-\beta(1-m_h^2)}, \end{aligned} \tag{4.7}$$

and the approximate Stein coupling is completed by taking $G:=(W'-W)/(2a)$. Below we work on (W,W',G,R), but note that all results easily transfer to $(\widetilde{W},\widetilde{W}',\widetilde{G},\widetilde{R})$ where $\widetilde{W}$ is as above, $\widetilde{W}'$ is defined in the obvious way, $\widetilde{G}=G/2$ and $\widetilde{R}'=R'/2$. In this case, $(\widetilde{W},\widetilde{W}')$ satisfy (1.6) and $|\widetilde{W}'-\widetilde{W}|\le 1$, so we apply our approximation framework, using Remark 2.2.

The first step is to bound the centred moments of $n^{-1/2}(W-nm_h)$. From [8], Proposition 1.3, for any fixed $k\ge 1$,

$$\mathbb{E}\left|\frac{W}{n}-\tanh\left(\beta\left(\frac{W}{n}\right)+h\right)\right|^k\le \mathrm{O}\big(n^{-k/2}\big). \tag{4.8}$$

Now, for y small enough, there exists $C'_y<\infty$ such that

$$\big|w-\tanh(\beta w+h)\big|\ge C'_y|w-m_h|\qquad \text{in } |w-m_h|\le y.$$

On the other hand, [14], Theorem 1.4, show that

$$\mathbb{P}\left(\left|\frac{W}{n}-m_h\right|>t\right)\le e^{-nC(t)}$$

for some $C(t)>0$, so that $\mathbb{P}(|n^{-1}W-m_h|>y)=\mathrm{O}(e^{-nC(y)})$. Combining these last two statements, it follows from (4.8) that, for any $k\ge 1$,

$$\mathbb{E}\left|\frac{W-nm_h}{\sqrt{n}}\right|^k=\mathrm{O}(1). \tag{4.9}$$

Now we turn to verifying (2.4); we use the representation in Remark 2.2. According to [34], Lemma 4.4, and using (4.9),

$$\begin{aligned} &\left|\mathbb{P}\big(W'-W=2|S\big)-\frac{(1-m_h)^2}{4}\right|\le Cn^{-1/2}\left(\frac{|W-\mu_n|}{\sigma_n}+n^{-1/2}\right),\\ &\left|\mathbb{P}\big(W'-W=2\big)-\frac{(1-m_h)^2}{4}\right|\le Cn^{-1/2}, \end{aligned}$$

so that (2.4) is satisfied for some constant $\kappa > 0$, with $k = 1$ and $T = 0$ almost surely:

$$\frac{1}{a}\bigl|\mathbb{P}\bigl(W' - W = 2|S\bigr) - \mathbb{P}\bigl(W' - W = 2\bigr)\bigr| \le \kappa\sigma_n\biggl(\frac{|W-\mu_n|}{\sigma_n} + 1\biggr).$$

We next show $\mathbb{E}[(R')^2] = \mathrm{O}(1)$. From (4.9) and (4.7),

$$\sqrt{\mathbb{E}\bigl[\bigl(R'\bigr)^2\bigr]} \le C\bigl(1 + |\mu_n - nm_h|\bigr).$$

To bound $|\mu_n - nm_h|$ when $h \ne 0$, note that the expectation of (4.4) is zero, which, with (4.5) and (4.6), implies that

$$|nm_h - \mu_n| \le \frac{|\sum_{i=1}^n \mathbb{E}(\tanh(\frac{\beta}{n}(W - S_i) + h) - \tanh(\frac{\beta}{n}W + h))| + Cn\mathbb{E}(\frac{W}{n} - m_h)^2}{|\beta(1 - m_h^2) - 1|};$$

applying (4.5) and (4.6) yields $|\mu_n - nm_h| = \mathrm{O}(1)$, and hence $\sqrt{\mathbb{E}[(R')^2]} = \mathrm{O}(1)$.

Collecting the results above, it now follows from Corollary 2.8(i) that

$$d_{\mathrm{TV}}\biggl(\mathcal{L}(\widetilde{W}), \mathrm{TP}\biggl(\frac{1}{2}\mu_n, \frac{1}{4}\sigma_n^2\biggr)\biggr) = \mathrm{O}\bigl(\sigma_n^{-1}\bigr) = \mathrm{O}\bigl(n^{-1/2}\bigr).$$

For the local limit bound, we only need to show

$$S_2\bigl(\mathcal{L}(\widetilde{W})\bigr) = \mathrm{O}\bigl(\sigma_n^{-2}\bigr),$$

which follows from [34], Lemma 4.4. Noting Remark 2.2, Corollary 2.8(ii) now easily implies that

$$d_{\mathrm{loc}}\biggl(\mathcal{L}(\widetilde{W}), \mathrm{TP}\biggl(\frac{1}{2}\mu_n, \frac{1}{4}\sigma_n^2\biggr)\biggr) = \mathrm{O}\bigl(\sigma_n^{-2}\bigr) = \mathrm{O}\bigl(n^{-1}\bigr).$$

Then μ_n can be replaced by nm_h and σ_n^2 by $\frac{n(1-m_h^2)}{(1-\beta+\beta m_h^2)}$. This follows from properties of the translated Poisson distribution, because $|\mu_n - nm_h| = \mathrm{O}(1)$, $|\sigma_n^2 - \frac{n(1-m_h^2)}{(1-\beta+\beta m_h^2)}| = \mathrm{O}(n^{1/2})$ and $n^{-1}\sigma_n^2$ is bounded away from 0.

Acknowledgements

ADB and NR thank the Institute of Mathematical Sciences and the Department of Statistics and Applied Probability at the National University of Singapore for their kind hospitality. AR was supported by NUS Research Grant R-155-000-167-112. ADB is supported in part by Australian Research Council Discovery Grants DP150101459 and DP150103588. NR is supported in part by Australian Research Council Discovery Grant DP150101459. We thank the two referees for their comments.

References

[1] Arratia, R. and Baxendale, P. (2015). Bounded size bias coupling: A Gamma function bound, and universal Dickman-function behavior. *Probab. Theory Related Fields* **162** 411–429. MR3383333

[2] Barbour, A.D. (1980). Equilibrium distributions for Markov population processes. *Adv. in Appl. Probab.* **12** 591–614. MR0578839

[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.

[5] Barbour, A.D. and Xia, A. (1999). Poisson perturbations. *ESAIM Probab. Stat.* **3** 131–150. MR1716120

[6] Bartroff, J., Goldstein, L. and Işlak, Ü. (2015). Bounded size biased couplings, log concave distributions and concentration of measure for occupancy models. *Bernoulli*. To appear. Available at arXiv:1402.6769v2.

[7] Bolthausen, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete* **66** 379–386. MR0751577

[8] Chatterjee, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072

[9] Chatterjee, S. and Dey, P.S. (2010). Applications of Stein's method for concentration inequalities. *Ann. Probab.* **38** 2443–2485. MR2683635

[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. and Fang, X. (2015). On the error bound in a combinatorial central limit theorem. *Bernoulli* **21** 335–359. MR3322321

[12] Chen, L.H.Y., Fang, X. and Shao, Q.-M. (2013). From Stein identities to moderate deviations. *Ann. Probab.* **41** 262–293. MR3059199

[13] Chen, L.H.Y. and Röllin, A. (2010). Stein couplings for normal approximation. Preprint. Available at arXiv:1003.6039v2.

[14] Dembo, A. and Montanari, A. (2010). Gibbs measures and phase transitions on sparse random graphs. *Braz. J. Probab. Stat.* **24** 137–211. MR2643563

[15] Eichelsbacher, P. and Löwe, M. (2010). Stein's method for dependent random variables occurring in statistical mechanics. *Electron. J. Probab.* **15** 962–988. MR2659754

[16] Ellis, R.S. (2006). *Entropy, Large Deviations, and Statistical Mechanics*. *Classics in Mathematics*. Berlin: Springer. Reprint of the 1985 original. MR2189669

[17] Ellis, R.S. and Newman, C.M. (1978). Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Z. Wahrsch. Verw. Gebiete* **44** 117–139. MR0503333

[18] 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

[19] Esseen, C.-G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law. *Acta Math.* **77** 1–125. MR0014626

[20] Fang, X. (2014). Discretized normal approximation by Stein's method. *Bernoulli* **20** 1404–1431. MR3217448

[21] Goldstein, L. (2005). Berry–Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.* **42** 661–683. MR2157512

[22] Goldstein, L. (2013). A Berry–Esseen bound with applications to vertex degree counts in the Erdős–Rényi random graph. *Ann. Appl. Probab.* **23** 617–636. MR3059270

[23] Goldstein, L. and Işlak, Ü. (2014). Concentration inequalities via zero bias couplings. *Statist. Probab. Lett.* **86** 17–23. MR3162712
[24] Goldstein, L. and Xia, A. (2006). Zero biasing and a discrete central limit theorem. *Ann. Probab.* **34** 1782–1806. MR2271482
[25] Hoeffding, W. (1951). A combinatorial central limit theorem. *Ann. Math. Stat.* **22** 558–566. MR0044058
[26] Kordecki, W. (1990). Normal approximation and isolated vertices in random graphs. In *Random Graphs '87* (*Poznań*, 1987) 131–139. Chichester: Wiley. MR1094128
[27] Krokowski, K., Reichenbachs, A. and Thäle, C. (2017). Discrete Malliavin–Stein method: Berry–Esseen bounds for random graphs and percolation. *Ann. Probab.* **45** 1071–1109. MR3630293
[28] McDonald, D.R. (1979). On local limit theorem for integer valued random variables. *Teor. Veroyatn. Primen.* **24** 607–614. MR0541375
[29] Petrov, V.V. (1975). *Sums of Independent Random Variables*. *Ergebnisse der Mathematik und ihrer Grenzgebiete* **82**. New York: Springer. Translated from the Russian by A.A. Brown. MR0388499
[30] Röllin, A. (2005). Approximation of sums of conditionally independent variables by the translated Poisson distribution. *Bernoulli* **11** 1115–1128. MR2189083
[31] Röllin, A. (2007). Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.* **17** 1596–1614. MR2358635
[32] Röllin, A. (2008). Symmetric and centered binomial approximation of sums of locally dependent random variables. *Electron. J. Probab.* **13** 756–776. MR2399295
[33] Röllin, A. (2017). Kolmogorov bounds for the normal approximation of the number of triangles in the Erdős–Rényi random graph. Preprint. Available at arXiv:1704.00410v1.
[34] Röllin, A. and Ross, N. (2015). Local limit theorems via Landau–Kolmogorov inequalities. *Bernoulli* **21** 851–880. MR3338649
[35] Ross, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132
[36] Stein, C. (1986). *Approximate Computation of Expectations*. *Institute of Mathematical Statistics Lecture Notes – Monograph Series* **7**. Hayward, CA: IMS. MR0882007
[37] Wald, A. and Wolfowitz, J. (1944). Statistical tests based on permutations of the observations. *Ann. Math. Stat.* **15** 358–372. MR0011424

Received July 2017 and revised November 2017