University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Barbour, AD;Ross, N;Wen, Y

Title:
Central moment inequalities using Stein's method

Date:
2020-01-01

Citation:
Barbour, A. D., Ross, N. & Wen, Y. (2020). Central moment inequalities using Stein's method. ELECTRONIC JOURNAL OF PROBABILITY, 25 (none), https://doi.org/10.1214/20-EJP493.

Persistent Link:
https://hdl.handle.net/11343/290322

License:
CC BY

Electron. J. Probab. **25** (2020), article no. 91, 1–21.
ISSN: 1083-6489 https://doi.org/10.1214/20-EJP493

Central moment inequalities using Stein's method

A. D. Barbour [*] Nathan Ross [†] Yuting Wen[‡]

Abstract

We derive explicit central moment inequalities for random variables that admit a Stein coupling, such as exchangeable pairs, size–bias couplings or local dependence, among others. The bounds are in terms of moments (not necessarily central) of variables in the Stein coupling, which are typically local in some sense, and therefore easier to bound. In cases where the Stein couplings have the kind of behaviour leading to good normal approximation, the central moments are closely bounded by those of a normal. We show how the bounds can be used to produce concentration inequalities, and compare them to those existing in related settings. Finally, we illustrate the power of the theory by bounding the central moments of sums of neighbourhood statistics in sparse Erdős–Rényi random graphs.

Keywords: moment inequalities; concentration inequalities; Stein's method; Erdős–Rényi random graph.
MSC2020 subject classifications: 60E15; 60C05.
Submitted to EJP on February 18, 2019, final version accepted on July 4, 2020.
Supersedes arXiv:1802.10225.

1 Introduction

Concentration inequalities are useful and powerful tools for estimating probabilities when exact computation is not possible. They have found important application in modern statistics [20]. Obtaining such inequalities has been an active area of probability for decades; for example, see [7] and the references there. The standard method for deriving strong concentration inequalities requires bounds on the moment generating function. For sums of independent random variables, this leads to Hoeffding/Bennett/Chernoff inequalities. Such results generalize to martingales in various ways [21], and to functions of independent random variables [5]. The research closest to this paper is concerned with showing concentration inequalities for random variables admitting various coupling

[*]Universität Zürich. http://user.math.uzh.ch/barbour/
E-mail: a.d.barbour@math.uzh.ch
[†]University of Melbourne. https://www.nathanrossprob.com
E-mail: nathan.ross@unimelb.edu.au
[‡]University of Melbourne.
E-mail: yutingyw@gmail.com

constructions related to Stein's method. The first results and key ideas are due to [8] and [10], who worked under the assumption that it is possible to construct an exchangeable pair that is marginally distributed as the variable of interest, and has certain conditional moment boundedness properties. The ideas were extended to bounded size–bias couplings by [15, 16], and then further to unbounded size–bias couplings satisfying a certain bounded in conditional probability assumption, by [14]. The exchangeable pair results have been generalized to the matrix setting by [19].

When it is not possible to bound the moment generating function, either due to its non-existence or to the complexity of the distribution of interest, another approach, that gives (sometimes only slightly) weaker concentration inequalities, is to bound moments of the variable of interest. For sums of independent random variables, there are a number of Rosenthal/Pinelis/Burkholder/Davis/Gundy inequalities. For functions of independent random variables, this program has been developed in [6]. Results for exchangeable pairs with additional structure are developed in [8, Theorem 1.5(iii)], with matrix versions in [22].

While Stein's method has found huge success in bounding the error made in distributional approximation (see [2], [12] and [9]), the techniques for concentration inequalities lag behind. This is essentially because controlling the moment generating function requires additional restrictive assumptions, such as boundedness, on the couplings that have found such great success in Stein's method for distributional approximation. In this paper, we develop explicit general bounds on the central moments of a random variable of interest, which are expressed in terms of the moments of the key variables appearing in a *Stein coupling*. In the context of sums of independent random variables, the analogues of these variables are the individual summands, and it is usual there to try to express a bound on the moments of the sum in terms of those of the summands. The Stein coupling formulation allows sums of random variables with a wide range of dependence structures to be treated in a unified way; the key variables may then be different, but their moments still provide accessible quantities for expressing a bound. We demonstrate the applicability of our approach in the settings of sums of independent random variables, size–bias couplings and local dependence, and we compare our bounds to those previously derived. However, the main advantage of our approach is that it can be used to obtain concentration in applications where other methods have not been developed. We illustrate this by showing concentration for the distribution of generic neighbourhood statistics in sparse Erdős–Rényi random graphs.

In the next section, we introduce the setting for the couplings we consider and state our moment inequalities. Section 3 contains a discussion on how these moment inequalities translate to concentration inequalities and Section 4 contains applications. The proofs are in Section 5.

2 Stein couplings and moment inequalities

The ordered collection of random variables (W, W', G, R) with $\mathbb{E}W = \mu$ are said to form an *approximate Stein coupling* if

$$\mathbb{E}[G(f(W') - f(W))] \;=\; \mathbb{E}[(W-\mu)f(W)] + \mathbb{E}[Rf(W)], \tag{2.1}$$

for all f such that the expectations exist. If the remainder R is identically zero, the triple (W, W', G) is called a *Stein coupling*. Standard examples of Stein couplings are discussed below in Section 4. These couplings are useful for establishing the error in approximating the distribution of W by a normal distribution, using Stein's method; see [11]. With $\sigma^2 := \operatorname{Var} W$, Stein's method shows that, if

$$|\,\mathbb{E}[(W-\mu)f(W) - \sigma^2 f'(W)]| \;\le\; \varepsilon(\sigma^2\|f'\|_\infty + \sigma^3\|f''\|_\infty)$$

for all f in a suitable class of functions with two bounded derivatives, then $d_{\mathrm{BW}}(\mathcal{L}(\sigma^{-1}(W-\mu)),\mathcal{N}(0,1)) = O(\varepsilon)$ as $\varepsilon \to 0$. Here, the bounded Wasserstein distance d_{BW} between probability measures P and Q on $\mathbb{R}$ is defined by

$$d_{\mathrm{BW}}(P,Q) \;:=\; \sup_{h\in\mathcal{H}_{\mathrm{BW}}} \Big| \int h\,dP - \int h\,dQ \Big|,$$

where $\mathcal{H}_{\mathrm{BW}} := \{h\colon \mathbb{R}\to\mathbb{R}\colon \|h\|_\infty \le 1, \|h'\|_\infty \le 1\}$; it metrizes weak convergence. Given a Stein coupling, and writing $D := W' - W$, it is immediate that

$$\begin{aligned}\mathbb{E}[(W-\mu)f(W) - \sigma^2 f'(W)] \;&=\; \mathbb{E}[G(f(W')-f(W))] - \sigma^2\,\mathbb{E}f'(W)\\ &=\; \mathbb{E}[(GD-\sigma^2)f'(W)] + \mathbb{E}[G(f(W+D)-f(W)-Df'(W))].\end{aligned}$$

Hence, to establish the accuracy of normal approximation, it suffices to show, for instance, that $\mathbb{E}|\,\mathbb{E}[GD\,|\,W] - \sigma^2| \le \varepsilon\sigma^2$ and that $\mathbb{E}|GD^2| \le \varepsilon\sigma^3$, for suitable choice of ε.

In this paper, we show that the existence of approximate Stein couplings also yields bounds for the even central moments of W. To some extent, the bounds cover cases where the distribution of W is not very close to being normal. To state our first theorem, we define $\|X\|_r := \{\mathbb{E}|X^r|\}^{1/r}$ for any random variable X and $r \in \mathbb{N} := \{1,2,\ldots\}$. Throughout the paper, we write $\mu := \mathbb{E}W$ and $\sigma^2 := \mathrm{Var}(W)$.

Theorem 2.1. *Let (W,W',G,R) be an approximate Stein coupling such that, for some constant ε such that $0 \le \varepsilon < 1$ and for some random variable $T \ge 0$,*

$$\big|\mathbb{E}[R|W]\big| \;\le\; \varepsilon|W-\mu| + T. \tag{2.2}$$

Let $k \in \mathbb{N}$ and suppose that $\|G\|_{2k} \le A_k =: A$ and that $\|D\|_{2k} \le B_k =: B$. Then

$$\begin{aligned}\|W-\mu\|_{2k} \;&\le\; \frac{A}{1-\varepsilon}\left[\left(1+\sqrt{\frac{B(1-\varepsilon)}{A(2k-1)}}\right)^{2k-1} - 1\right] + \frac{\|T\|_{2k}}{1-\varepsilon}\\ &\le\; \sqrt{\frac{(2k-1)AB}{1-\varepsilon}}\exp\left\{\sqrt{\frac{B(1-\varepsilon)(2k-1)}{A}}\right\} + \frac{\|T\|_{2k}}{1-\varepsilon}\,.\end{aligned}$$

The statement of the theorem looks at first sight rather complicated. Before discussing it, we state a related result with a slightly simpler bound under the additional assumption that, for each $r \in \mathbb{N}$ of interest, $\mathbb{E}\{|W'-\mu|^r\} \le \mathbb{E}\{|W-\mu|^r\}$, together with an explicit statement about the smallness of $\|T\|_r$. The extra assumption is satisfied for exchangeable pairs, when $\mathcal{L}(W') = \mathcal{L}(W)$, as well as for the Stein coupling that we shall use for sums of independent random variables with mean zero. Note that, in this theorem, odd powers of $|W-\mu|$ are also allowed.

Theorem 2.2. *Let (W,W',G,R) be an approximate Stein coupling such that (2.2) is satisfied. Then, for any $r \in \mathbb{N}$ such that $\mathbb{E}\{|W'-\mu|^r\} \le \mathbb{E}\{|W-\mu|^r\}$ and that $\sigma^{-1}\|T\|_r \le \varepsilon'$ for some $\varepsilon' < 1-\varepsilon$, it follows that*

$$\|W-\mu\|_r \;\le\; \sqrt{\frac{2(r-1)\|G\|_r\|D\|_r}{1-\varepsilon-\varepsilon'}}\,.$$

Theorems 2.1 and 2.2 provide bounds for central moments of W, expressed simply in terms of the moments of the random variables G and D, which, in practical circumstances, are easier to handle than W itself. However, it is not at first sight clear how good the bounds are likely to be. To get some idea of this, consider the case of a Stein coupling,

when $R = 0$. In this case, taking $f(w) = w - \mu$ in (2.1), it follows that $\sigma^2 = \mathbb{E}(GD)$. Thus, using the Cauchy–Schwarz and Hölder inequalities, we have

$$\sigma \;=\; \{\mathbb{E}(GD)\}^{1/2} \;\le\; \{\|G\|_2\|D\|_2\}^{1/2} \;\le\; \{\|G\|_{2k}\|D\|_{2k}\}^{1/2} \;\le\; \{AB\}^{1/2}.$$

Hence Theorems 2.1 and 2.2 show that $\|W - \mu\|_{2k}$ is sandwiched between σ and a multiple of $\{AB\}^{1/2}$, which is the upper bound for σ obtained by replacing the 2-norms of G and D in the Cauchy–Schwarz inequality by their $2k$-norms. In many applications concerning asymptotics as the size n of a problem increases, the distribution of D remains more or less constant, while σ^2 grows like n. Since $\mathbb{E}(GD) = \sigma^2$, the distribution of $n^{-1}G$ also remains more or less constant. Hence B can typically be chosen more or less constant in n, whereas A is proportional to n, implying that $B/A = O(n^{-1})$. Thus the exponential factor in the second inequality of Theorem 2.1 is close to 1 unless k is very large, making the constant multiplying $\{AB\}^{1/2}$ smaller than that in Theorem 2.2.

For example, for a sum $W := \sum_{i=1}^n X_i$ of independent and identically distributed random variables with zero mean, there is a Stein coupling $(W, W - X_I, -nX_I)$, where I denotes a random variable with the uniform distribution on $\{1, 2, \dots, n\}$ that is independent of $(X_i)_{i=1}^n$, so that, in particular, $\varepsilon = 0$. Then $\sigma^2 = n\,\mathbb{E}(X_1^2)$ and G have factors of n, whereas $\mathbb{E}(D^2) = \mathbb{E}(X_1^2)$ is constant in n. In particular, we can take A and B such that $AB = n\|X_1\|_{2k}^2$, as compared to $\sigma^2 = n\|X_1\|_2^2$, and, for fixed k, the bound on $\|W - \mu\|_{2k}$ in Theorem 2.1 is close to $\sqrt{n(2k-1)}\|X_1\|_{2k}$. This Stein coupling satisfies the extra condition of Theorem 2.2, giving the bound $\sqrt{2n(2k-1)}\|X_1\|_{2k}$. Note that, by considering $X_i \sim 2I - 1$, where $I \sim \text{Bernoulli}(1/2)$, the factor of $\sqrt{2k-1}$ is seen to be inevitable, being the rate of the growth of $\|N\|_{2k}$ for N standard normal; see Lemma 2.5.

Theorems 2.1 and 2.2 are our most useful results for difficult applications, such as Theorem 4.2 below. However, we also show that sharper bounds can be obtained, if Stein's method of normal approximation applies to the variable under consideration. The next theorem shows that, if this is the case, the product AB in the leading term can be replaced by σ^2. The extra condition, given in (2.4), would be expected to be satisfied with ε_3 small, if normal approximation were good. Thus, if normal approximation is good enough, the even central moments of W cannot be too much bigger than those of the corresponding normal distribution.

Theorem 2.3. *With the assumptions and notation of Theorem 2.1, suppose that now, for non-negative random variables T_1 and T_2 and for non-negative ε_i, $1 \le i \le 3$, we have*

$$\big|\mathbb{E}[R|W]\big| \;\le\; \varepsilon_1|W - \mu| + T_1, \tag{2.3}$$

where $\sigma^{-1}\|T_1\|_{2k} \le \varepsilon_2$, and

$$|\,\mathbb{E}[GD\,|\,W] - \sigma^2| \;\le\; \varepsilon_3(W - \mu)^2 + T_2. \tag{2.4}$$

Then, if $\varepsilon_1 + \varepsilon_2 + (2k-1)\varepsilon_3 < 1$, it follows that

$$\|W - \mu\|_{2k} \;\le\; \frac{\sigma\sqrt{2k-1}}{\sqrt{1 - \varepsilon_1 - \varepsilon_2 - (2k-1)\varepsilon_3}}\left(1 + \frac{(k-1)AB^2}{\sigma^3\sqrt{2k-1}}\exp\Big\{\frac{B\sqrt{2k-1}}{\sigma}\Big\} + \frac{\|T_2\|_k}{\sigma^2}\right)^{1/2}.$$

Note that it is always possible to take $\varepsilon_3 = 0$ and $T_2 := |\,\mathbb{E}[GD\,|\,W] - \sigma^2|$. Note also that the conditions of Theorem 2.3 are similar to others used in the Stein's method literature; for example, see [13, Theorem 3.1], [3, (2.4)], and [24, Theorem 3.11], covering applications such as the anti-voter model and the number of isolated vertices in a sparse Erdős–Rényi random graph.

Supposing that $\sigma^{-2}\|T_2\|_k \le \varepsilon_4$ with ε_4 small enough, we have a further refinement, showing that the even central moments of W are bounded by a quantity which, under

typical asymptotics, is equivalent to the corresponding normal moment. To state the theorem, we define

$$h_k \;:=\; 2^{-1/2}e^{5/2}\sigma^{-3}AB^2\sqrt{k-1}. \tag{2.5}$$

Theorem 2.4. *Under the assumptions of Theorem 2.3, and assuming that $\sigma^{-2}\|T_2\|_k \le \varepsilon_4$, suppose that*

$$E \;:=\; \varepsilon_1+\varepsilon_2+(2k-1)(\varepsilon_3+\varepsilon_4) \;<\; 1-h_k,$$

and that $\sigma \ge B\sqrt{e(2k-1)}$. Then

$$\|\sigma^{-1}(W-\mu)\|_{2k} \;\le\; \|N\|_{2k}/\sqrt{1-E-h_k}.$$

where N denotes a standard normal random variable.

We end this section with a lemma that is used in our proofs, but is also important in interpreting the sharpness of central moment bounds.

Lemma 2.5. *For N a standard normal random variable, define*

$$c_1(k) \;:=\; \{\sqrt{2k-1}/\|N\|_{2k}\}.$$

Then c_1 is increasing in k, with $c_1(1)=1$ and $c_1(\infty)=\sqrt{e}$.

Proof. First, $\log c_1(k) < 1/2$ follows using the usual bounds for the error in Stirling's formula. Then, using the expansion of $(x+1/2)\log(1+1/x)$ for $x>1$ in the first inequality, we find that

$$\frac{c_1(k+1)^{2(k+1)}}{c_1(k)^{2k}} \;=\; \Big(1+\frac{2}{2k-1}\Big)^k \;>\; e \;>\; c_1(k)^2, \tag{2.6}$$

which implies that $c_1(k+1)>c_1(k)$, and then also that $\lim_{k\to\infty} c_1(k)=\sqrt{e}$. □

3 Concentration inequalities from central moment bounds

The bounds derived above can be used with Markov's inequality to show that the distribution of a random variable W is concentrated about its mean, by starting from

$$\mathbb{P}[|W-\mu|>t] \;\le\; t^{-2k}\|W-\mu\|_{2k}^{2k}, \tag{3.1}$$

and choosing k carefully. Suppose that $(W_n,\, n\ge 1)$ is a sequence of random variables with means μ_n and variances $\sigma_n^2 \asymp n$. Then one weak form of concentration is to say that W_n is concentrated about its mean μ_n on the scale $(d_n)_{n\ge1}$ if, for any $r\ge1$, there exist $K(r)$ and $c(r)$ such that

$$\mathbb{P}[|W_n-\mu_n|>c(r)d_n] \;\le\; K(r)n^{-r},$$

uniformly in n. Under the assumptions of Theorem 2.1, and for an exact Stein coupling with $\|n^{-1}G\|_{2k}\le\alpha_{n,k}$ and $\|D\|_{2k}\le\beta_{n,k}$, we have,

$$\mathbb{P}[|W_n-\mu_n|>t_n] \;\le\; t_n^{-2k}\Big\{n(2k-1)\alpha_{n,k}\beta_{n,k}\Big\}^k \exp\Big\{\frac{2k}{\sqrt{n}}\sqrt{\frac{(2k-1)\beta_{n,k}}{\alpha_{n,k}}}\Big\}. \tag{3.2}$$

Thus, taking $k=k_n:=\log n$ and

$$d_n \;:=\; \sqrt{n(2k_n-1)\alpha_{n,k_n}\beta_{n,k_n}}\exp\Big\{\frac{1}{\sqrt{n}}\sqrt{\frac{(2k_n-1)\beta_{n,k_n}}{\alpha_{n,k_n}}}\Big\}, \tag{3.3}$$

we have

$$\mathbb{P}[|W_n-\mu_n|>cd_n] \;\le\; c^{-2\log n} \;=\; n^{-2\log c},$$

which can be made smaller than n^{-r}, for any given r, by choosing $c = c(r)$ large enough. Thus, in this sense, Theorem 2.1 shows that W_n is concentrated around μ_n on a scale $(d_n)_{n\geq 1}$, where d_n is as in (3.3).

As remarked earlier, it is often the case in such asymptotics that the distributions of D and $n^{-1}G$ are more or less constant in n, in the sense that their tails are uniformly bounded in n. However, for any n, the norms $\|D\|_k$ and $\|n^{-1}G\|_k$ grow to infinity as k increases, unless the random variables themselves are uniformly bounded. The following lemma is useful in determining how fast the norms grow with k; its proof is straightforward, by using, for example, saddle point methods.

Lemma 3.1. *Let X be a random variable such that, for some $a, b, c > 0$,*

$$\mathbb{P}[|X| > x] \;\leq\; ce^{-bx^a} \quad \textit{for all } x > 0. \tag{3.4}$$

Then, for $k \geq 1$,

$$\mathbb{E}|X|^{2k} \;\leq\; \frac{2kc}{a}\left(b^{-1/a}\right)^{2k}\Gamma(2k/a).$$

If, instead, $a > 1$ and

$$\mathbb{P}[|X| > x] \leq ce^{-b(\log(x+1))^a} \quad \textit{for all } x > 0, \tag{3.5}$$

then, for $k \geq 1$,

$$\mathbb{E}|X|^{2k} \;\leq\; cC_1(a,b)k^{a/(2(a-1))}\exp\{C_2(a,b)k^{a/(a-1)}\},$$

for suitable constants $C_1(a,b)$ and $C_2(a,b)$.

Hence the quantity d_n above is of order $O(\sqrt{n\log n})$ if both D and $n^{-1}G$ are a.s. bounded for all n by the same constants x_1 and x_2. If both D and $n^{-1}G$ have tails bounded as in (3.4), uniformly for all n, then $d_n = O((\log n)^{1/a+1/2}\sqrt{n})$; if they have tails bounded as in (3.5), uniformly for all n, and if $a > 2$, then $d_n = O(n^{1/2+\delta})$ for any $\delta > 0$.

The classical large deviation bounds, such as the Chernoff bounds, deliver much smaller bounds for the probabilities of large deviations than those required for the weaker form of concentration discussed above. We now show that our moment bounds can also deliver analogous results, if values of k_n larger than $\log n$ are taken.

Under the conditions of Theorem 2.1, we can invoke (3.2), and choose $k = k_n(t)$ to make the principal factor $t^{-2k}\Big\{n(2k-1)\|n^{-1}G\|_{2k}\|D\|_{2k}\Big\}^k$ small, for fixed choice of t. In particular we have the following corollary for a.s. bounded $n^{-1}G$ and D.

Corollary 3.2. *Under the assumptions of Theorem 2.1, for an exact Stein coupling such that $|D| \leq x_1$ and $|n^{-1}G| \leq x_2$ a.s. for all n, we have, for any $t \geq \sqrt{2nx_1x_2e}$,*

$$\mathbb{P}[|W_n - \mu_n| > t] \;\leq\; e\exp\left\{-\frac{t^2}{2nx_1x_2e}\left(1 - \frac{2t}{x_2n\sqrt{e}}\right)\right\}. \tag{3.6}$$

The corollary gives good bounds as long as $t \ll n$. However, the factor e in the denominator makes the exponent smaller than that in the Chernoff bound; we return to this later.

If we only have control over the tails of $|D|$ and $|n^{-1}G|$, we can obtain the following.

Corollary 3.3. *Under the assumptions of Theorem 2.1, for an exact Stein coupling such that both $|D|$ and $|n^{-1}G|$ have tails bounded as in* (3.4)*, then for $t^2 \geq ne2^{(a+2)/a}(ab)^{-2/a}$, we have*

$$\mathbb{P}[|W_n - \mu_n| > t] \;\leq\; c\sqrt{2\pi}\left\{\frac{2k_t}{a} + \frac{1}{2}\right\}^{1/2} e^{-k_t(a+2)/a}\exp\{n^{-1/2}(2k_t)^{3/2}\},$$

where $k_t := \left\lfloor \frac{1}{2}\big(t^2/(ne)\big)^{a/(a+2)}(ab)^{2/(a+2)}\right\rfloor$.

It is clear that under the hypotheses of the corollary, there are constants K_1, K_2 and K_3 depending on a and b such that

$$\mathbb{P}[|W_n - \mu_n| > t] \;\le\; cK_1(t^2/n)^{a/(2(a+2))}\exp\Big\{-K_2\Big(\frac{t^2}{n}\Big)^{a/(a+2)}\Big\{1 - K_3\frac{t}{n}\Big\}\Big\}.$$

The *power* of t in the exponent is now no longer as large as 2, though it approaches 2 as a increases. On the other hand, the bound still gives useful results, provided that $t \ll n$.

If the conditions of Theorem 2.4 are satisfied, better bounds can be obtained; in particular, if we have an exact Stein coupling for which the k-norms of $T_2 := |\,\mathbb{E}[GD\,|\,W] - \sigma^2|$ can be shown to be suitably small.

Corollary 3.4. *Under the assumptions of Theorem 2.4, taking $k = k(y) := \lceil y^2/2\rceil$, we have*

$$\mathbb{P}[\sigma_n^{-1}|W_n - \mu_n| > y] \;\le\; e\sqrt{2}\exp\Big\{-\frac{y^2}{2}\Big\}(1 - E - h_k)^{-k}. \tag{3.7}$$

For an exact Stein coupling, the quantities ε_1 and ε_2 are zero. In most applications they, and ε_3 and ε_4, can be expected to depend on n as a power $n^{-\alpha}$, for a suitable $\alpha > 0$ (often $\alpha = 1/2$). However, ε_2 and ε_4 also involve k-norms of the error random variables T_1 and T_2, and the quantity h_k also involves k-norms of D and $n^{-1}G$. Assuming that their tails are bounded as in (3.4), these norms can be dealt with much as above, resulting in powers of k as factors; so long as $y \ll n^\beta$ for a suitably small index β, the bound above is then useful. In particular, if $|D|$, $|n^{-1}G|$, $n^{-1/2}|T_1|$ and $n^{-1}|T_2|$ are uniformly bounded, then good bounds are obtained for $y \ll n^{1/4}$, equivalent to deviations of W_n from its mean of order $o(n^{3/4})$. This is not as good a range as for the Chernoff bounds, but the main exponent is ideal.

4 Applications

Here we use Theorems 2.1 – 2.4 in some applications.

4.1 Sums of independent random variables

Let $W := \sum_{i=1}^n X_i$, where $X_1, \dots, X_n$ are independent mean zero random variables with $\mathbb{E}|X_i|^{2k} < \infty$ for $1 \le i \le n$. Let $W' = W - X_I$, where I is a uniform index from $\{1, \dots, n\}$, and is independent of $(X_i)_{i=1}^n$. Then for $G := -nX_I$, as mentioned above, it is easily checked that (W, W', G) is a Stein coupling; see [11]. Note too that, for $D := W' - W = -X_I$, we have $G = nD$ and so we can take $A = nB$, with

$$B \;=\; \|X_I\|_{2k} \;=\; \Big\{n^{-1}\sum_{i=1}^n \mathbb{E}X_i^{2k}\Big\}^{1/2k}.$$

Writing $\rho_k := \|X_I\|_{2k}/\|X_I\|_2$, the bound from Theorem 2.1 becomes

$$\begin{aligned}\|\sigma^{-1}W\|_{2k} \;&\le\; \frac{n\|X_I\|_{2k}}{\sigma}\left[\left(1 + \sqrt{\frac{1}{n(2k-1)}}\right)^{2k-1} - 1\right]\\ &\le\; \rho_k\sqrt{(2k-1)}\exp\left\{\sqrt{\frac{(2k-1)}{n}}\right\}.\end{aligned}$$

The theorems that give sharper bounds rely on establishing (2.4). For the coupling above, writing $\sigma_i^2 := \mathbb{E}X_i^2$, the simplest version is obtained by taking $\varepsilon_3 = 0$ and

$$T_2 \;:=\; \mathbb{E}\Big\{\sum_{i=1}^n (X_i^2 - \sigma_i^2)\,|\,W\Big\},$$

but this yields a result that is not as clean as the one following, that we derive by a slightly different argument.

Proposition 4.1. *For $W := \sum_{i=1}^n X_i$, where $X_1, \dots, X_n$ are independent mean zero random variables with $\mathbb{E}|X_i|^{2k} < \infty$ for $1 \le i \le n$ and an integer $k \ge 2$, we have*

$$\|\sigma^{-1}W\|_{2k} \;\le\; \|N\|_{2k}/\sqrt{1-h_k'}, \tag{4.1}$$

where

$$h_k' \;:=\; 5\sqrt{2e^3}\rho_k^3\sqrt{\frac{k-1}{n}} \quad \textit{and} \quad \rho_k \;:=\; \|X_I\|_{2k}/\|X_I\|_2,$$

and n is assumed to be large enough that $h_k' < 1$.

To interpret the proposition, note that, if there is uniform control over the tails of the X_i, such as in Lemma 3.1, then $h_k' \to 0$ as $n \to \infty$ for any fixed k. Furthermore, even for tails as given by (3.5), $h_k' \to 0$ for k growing as $(\alpha \log n)^\beta$, for $0 < \beta < a - 1$.

4.2 Local dependence

Let $W := \sum_{i=1}^n X_i$, where the random variables $X_1, \dots, X_n$ have mean zero, and where $\mathbb{E}|X_i|^{2k} < \infty$, for $1 \le i \le n$. Assume also, for each $i = 1, \dots, n$, that there are "neighbourhoods" $\mathcal{N}_i \subseteq \{1, \dots, n\}$ such that X_i is independent of $W_i := W - \sum_{j \in \mathcal{N}_i} X_j$. Let $W' := W_I$, where I is a uniform index from $\{1, \dots, n\}$, independent of all else, and set $G := -nX_I$. Then, it is easily checked that (W, W', G) is an exact Stein coupling; see also [11, Construction 2A]. Here $D := W' - W = -\sum_{j \in \mathcal{N}_I} X_j$ Thus we can apply the bound of Theorem 2.1 with

$$A \;=\; n\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\big|X_i\big|^{2k}\right)^{1/(2k)}; \qquad B \;=\; \left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}\Big|\sum_{j \in \mathcal{N}_i} X_j\Big|^{2k}\right)^{1/(2k)}.$$

If, for example, for all $i = 1, \dots, n$, $d := \max_i |\mathcal{N}_i|$ and $x := \max_i \|X_i\|_{2k}$, then, using Minkowski's inequality, we could also take

$$A \;=\; nx, \;\; B \;=\; dx,$$

and Theorem 2.1 implies that

$$\begin{aligned} \|W\|_{2k} \;&\le\; nx\left[\left(1 + \sqrt{\frac{d}{n(2k-1)}}\right)^{2k-1} - 1\right] \\ &\le\; \sqrt{n}x\sqrt{d(2k-1)}\exp\left\{\sqrt{\frac{d(2k-1)}{n}}\right\}. \end{aligned} \tag{4.2}$$

Corollary 3.2 can also be invoked, giving a bound on large deviation probabilities of

$$\mathbb{P}[|W| > t] \;\le\; e\exp\left\{-\frac{t^2}{2endx^2}\left(1 - \frac{t}{nx}\sqrt{\frac{6}{e}}\right)\right\}, \tag{4.3}$$

for all $t \ge \sqrt{2endx^2}$. This can be compared to [18, Theorem 2.1], which, under the stronger condition that $|X_i| \le x$, gives

$$\mathbb{P}[|W| > t] \;\le\; 2\exp\left\{-\frac{t^2}{2ndx^2}\right\}. \tag{4.4}$$

This bound is better than (4.3) because it does not have the factor of e in the denominator of the exponent. However, in situations where the refined inequalities such as Corollary 3.4 apply, our results improve on (4.4). Scan statistics furnish standard examples of this kind. If $Y_1, Y_2, \dots, Y_n$ are independent random variables, and, for $j = 1, \dots, n-\ell+1$, we define

$$X_j \;:=\; \varphi_j(Y_j, \dots, Y_{j+\ell-1}) - \mathbb{E}\varphi_j(Y_j, \dots, Y_{j+\ell-1}),$$

where $\varphi_j \colon \mathbb{R}^\ell \to \mathbb{R}$ is uniformly bounded for all j, then $W = \sum_{j=1}^{n-\ell+1} X_j$ satisfies the hypotheses above with $d = 2\ell + 1$. The special case of head runs in Bernoulli trials is an example in the next section, where we show how to improve the basic inequality.

4.3 Size–bias couplings

For a random variable $W \geq 0$ with $\mathbb{E}W = \mu < \infty$, we say that W^s has the size–bias distribution of W if

$$\mathbb{E}f(W^s) \;=\; \frac{\mathbb{E}[Wf(W)]}{\mu},$$

for all functions f such that the right hand side is well defined. If (W, W^s) is a coupling of a distribution with its size–bias distribution, and we define $W' := W^s$ and $G := \mu$, then (W, W', G) is a Stein coupling. Theorem 2.1 easily applies in this setting with $D := W^s - W$, so that we can take

$$A \;=\; \mu, \qquad B \;=\; \bigl(\mathbb{E}\bigl[|W^s - W|^{2k}\bigr]\bigr)^{1/(2k)}.$$

For *bounded* size–bias couplings, where $|W^s - W| \leq c$ for some constant c, then $A = \mu$ and $B \leq c$, and in this case the bound (3.6) becomes

$$\mathbb{P}[|W - \mu| > t] \;\leq\; e \exp\biggl\{-\frac{t^2}{2e\mu c}\biggl(1 - \frac{t}{\mu}\sqrt{\frac{6}{e}}\biggr)\biggr\}. \tag{4.5}$$

This is to be compared with the best known bounds under these hypotheses, those of [1]. A very slightly weaker version of their bound is

$$\mathbb{P}[|W - \mu| > t] \;\leq\; 2 \exp\biggl\{-\frac{t^2}{2\mu c + 2ct/3}\biggr\}. \tag{4.6}$$

This is better than (4.5), because it does not have the factor e in the denominator of the exponent.

However, in many circumstances, Theorem 2.4 can also be applied, and Corollary 3.4 then yields bounds for large deviation probabilities. The quantity h_k defined in (2.5) is already directly expressed in terms of A and B. Because the size–bias coupling yields an exact Stein coupling, $\varepsilon_1 = \varepsilon_2 = 0$. We can also take $\varepsilon_3 = 0$, in which case

$$\sigma^{-2} T_2 \;=\; \frac{\mu}{\sigma^2}\Bigl|\mathbb{E}\{D \,|\, W\} - \frac{\sigma^2}{\mu}\Bigr| \;=\; |\mathbb{E}\{D \,|\, W\} - \mathbb{E}D| / \mathbb{E}D.$$

Thus, in order to use (3.7), all that we need in addition is a useful bound on the k-norm $\| \mathbb{E}\{D \,|\, W\} - \mathbb{E}D\|_k$. In practice, it may be easier to bound $\| \mathbb{E}\{D \,|\, \mathcal{F}\} - \mathbb{E}D\|_k$ for a larger σ-field $\mathcal{F}$, with respect to which W is measurable. Provided this norm is sufficiently small, Corollary 3.4 yields bounds in which the leading term in the exponent is improved to $-t^2/(2\sigma^2)$, though typically in a restricted range of t; note that the leading term in the exponent is now typically better than $-t^2/(2\mu c)$, since $\sigma^2 = \mu\,\mathbb{E}D$, and $\mathbb{E}D \leq c$, with equality only if $\mathcal{L}(a_1 W) = \mathrm{Po}(a_2)$, for some $a_1, a_2 > 0$.

For example, [16] derive a bound of essentially the same form as (4.6) for the number of m-runs in n i.i.d. Bernoulli trials $\xi_1, \dots, \xi_n$, with common success probability p, in

which case $c = 2m-1$. Letting $\mathcal{F} := \{\xi_1, \dots, \xi_n\}$, $\mathbb{E}\{D \,|\, \mathcal{F}\}$ can be expressed as an average of n non-negative locally dependent random variables, each bounded above by $2m-1$, with dependence neighbourhoods of size $3m-2$; its $2k$-norm can then be bounded using (4.2) by $5em^{3/2}\sqrt{k/n}$, if $6km \le n$. Thus

$$(1 - E - h_k)^{-k} \;\approx\; 1 \qquad \text{if} \quad k^{5/2} \;\ll\; n^{1/2}\,\mathbb{E}D,$$

so that the bound in (3.7) can be used as long as $y \ll n^{1/10}$, if $\mathbb{E}D$ is bounded away from zero (or, equivalently, if p is bounded away from 1). This yields a leading term in the exponent of $-t^2/(2\mu\,\mathbb{E}D)$, for $t \ll \sigma n^{1/10}$, rather than $-t^2/(2\mu(2m-1))$, with

$$1 - p^m \;\le\; \mathbb{E}D \;=\; 1 + 2\,\frac{p - p^m}{1-p} - (2m-1)p^m \;\le\; 7(2m-1)/12,$$

for all p and $m \ge 2$, and indeed with $8/9 \le \mathbb{E}D \le 2$ for all $p \le 1/3$ and $m \ge 2$. This represents a considerable improvement over (4.6) in the given range of t.

Thus we see that our approach can compete with best known bounds, in situations where other bounds are available. However, we emphasize again that the main advantage of our approach is that it gives bounds in many settings where other methods cannot be applied. For size–bias couplings in which the random variable $|W^s - W|$ is not uniformly bounded, for instance, exponential concentration inequalities are difficult to come by; see [17].

4.4 Local neighbourhood statistics of Erdős–Rényi random graphs

Let $\mathbb{G}_n$ be the set of simple and undirected graphs on n vertices with labels $[n] := \{1, \dots, n\}$, and let $\mathcal{G}_n$ be an Erdős–Rényi random graph on $\mathbb{G}_n$, with edge probability $p := \lambda/n$. Fix $r \in \mathbb{N}$ and, for $G \in \mathbb{G}_n$ and each $i = 1, \dots, n$, let $\mathcal{N}_r(i, G)$ be the "r-neighbourhood" consisting of the subgraph induced by all vertices at a distance no greater than r from vertex i in G, with vertex i distinguished; note that this includes the edges between vertices at graph distance r from vertex i. For each $i = 1, \dots, n$, let U be a real-valued function on graphs on at most n vertices having a distinguished vertex, and set

$$X_i \;:=\; U\big(\mathcal{N}_r(i, \mathcal{G}_n)\big); \qquad W \;:=\; \sum_{i=1}^{n} X_i.$$

Concrete examples are given by taking U to be any function of the degree of the distinguished vertex, with $r = 1$, or the number of copies of some fixed subgraph H containing the distinguished vertex, with $r = \mathrm{diameter}(H)$. In the latter case, W equals the number of occurrences of H in the graph, times the number of vertices of H.

For sparse Erdős–Rényi random graphs, r-neighbourhoods are small with high probability. Hence, as long as U is well behaved, $\|D\|_{2k}$ should be of a good order. We illustrate this principle in the following theorem. For a graph G, let $V(G)$ denote its vertex set, and $|V(G)|$ the number of its vertices.

Theorem 4.2. *Fix $r \in \mathbb{N}$ and $\beta \ge 0$, and let U be a function on graphs with a distinguished vertex, as above, such that, for some positive constant c and any graph G with a distinguished vertex in its domain,*

$$|U(G)| \;\le\; c|V(G)|^{\beta}. \tag{4.7}$$

Fix $\lambda > 0$ and let $\mathcal{G}_n$ be an Erdős–Rényi random graph on $\mathbb{G}_n$, with edge probability $p := \lambda/n$, and define

$$X_i \;:=\; U\big(\mathcal{N}_r(i, \mathcal{G}_n)\big)$$

and $W = \sum_{i=1}^n X_i$. *Then*

$$n^{-1/2}\|W - \mathbb{E}W\|_q \;\le\; cC(r,\beta)\max\{\lambda, q(1+\beta)\}^{(1+2\beta)r+1/2},$$

where

$$C(r,\beta) = \sqrt{2\big(20^{1+\beta} + 4^{1+\beta}\big)}\left(\frac{\pi e^{e-2}}{\log(e-1)}\right)^{(1+2\beta r)}.$$

For some basic examples, if U is the indicator that the degree of the distinguished vertex i is in some set, we set $r = 1$ and can take $c = 1$ and $\beta = 0$ in (4.7). If U is the number of copies of a fixed subgraph H containing i, we set $r = \mathrm{diameter}(H)$ and can take $c = |V(H)|$ and $\beta = |V(H)| - 1$. These examples are standard, and may also be handled by other methods. However, the theorem applies to more exotic statistics, such as the number of vertices having at least degree d with at most k neighbours having degree no greater than d ($r = 2$, $c = 1$, $\beta = 0$).

If X is a random variable such that $\|X - \mathbb{E}X\|_k \le (Ck)^\alpha$ for all $k \ge k_0$, then it follows from Markov's inequality that

$$\mathbb{P}[|X - \mathbb{E}X| \ge t] \;\le\; \{(Ck)^\alpha/t\}^k, \quad k \ge k_0.$$

Taking $k = C^{-1}(t/e)^{1/\alpha}$, this gives

$$\mathbb{P}[|X - \mathbb{E}X| \ge t] \;\le\; e^{-C^{-1}(t/e)^{1/\alpha}}, \quad t \ge e(Ck_0)^\alpha.$$

Here, we take $X := W/\{cC(r,\beta)\sqrt{n}\}$, $C := (1+\beta)$, $k_0 := \lambda/(1+\beta)$ and $\alpha := (1+2\beta)r+1/2$. In particular, if $n \to \infty$ and r, U and λ remain the same, then $W_n - \mathbb{E}W_n$ is weakly concentrated on the scale $\sqrt{n}\{\log n\}^{(1+2\beta)r+1/2}$.

5 Proofs

In this section we prove the previous results.

5.1 Proofs for general inequalities

Proof of Theorem 2.1. For $f(w) = (w-\mu)^{2k-1}$, using (2.1), we have

$$\mathbb{E}[G\{f(D+W) - f(W)\}] \;=\; \mathbb{E}\Big[(W-\mu)^{2k}\Big] + \mathbb{E}\Big[R(W-\mu)^{2k-1}\Big].$$

Hence

$$(1-\varepsilon)\,\mathbb{E}\Big[(W-\mu)^{2k}\Big] \;\le\; |\mathbb{E}[G\{f(D+W) - f(W)\}]| + \mathbb{E}\Big[T\big|W-\mu\big|^{2k-1}\Big].$$

Furthermore, by the binomial theorem, we have

$$f(D+W) - f(W) \;=\; \sum_{j=1}^{2k-1}\binom{2k-1}{j} D^j (W-\mu)^{2k-1-j}.$$

Now, setting $x := \|W-\mu\|_{2k}^2$ and using the triangle inequality, it follows that

$$(1-\varepsilon)x^k \;\le\; \sum_{j=1}^{2k-1}\binom{2k-1}{j}\mathbb{E}\big[\big|GD^j(W-\mu)^{2k-j-1}\big|\big] + \mathbb{E}\Big[T\big|W-\mu\big|^{2k-1}\Big]. \tag{5.1}$$

Using Hölder's inequality, and because $\|G\|_{2k} \le A$ and $\|D\|_{2k} \le B$, this gives

$$\begin{aligned}(1-\varepsilon)x^k \;&\le\; A\sum_{j=1}^{2k-1}\binom{2k-1}{j}B^j x^{\frac{2k-j-1}{2}} + \|T\|_{2k}x^{k-1/2}\\ &=\; x^{k-\frac{1}{2}}\left[A\sum_{j=1}^{2k-1}\binom{2k-1}{j}\left(\frac{B}{\sqrt{x}}\right)^j + \|T\|_{2k}\right].\end{aligned}$$

Rearranging, we thus have

$$\begin{aligned}\sqrt{x} \;&\leq\; \frac{A}{1-\varepsilon}\sum_{j=1}^{2k-1}\binom{2k-1}{j}\Big(\frac{B}{\sqrt{x}}\Big)^{j} + \frac{\|T\|_{2k}}{1-\varepsilon}\\ &=\; \frac{A}{1-\varepsilon}\left[\Big(1+\Big(\frac{B}{\sqrt{x}}\Big)\Big)^{2k-1} - 1\right] + \frac{\|T\|_{2k}}{1-\varepsilon}.\end{aligned} \tag{5.2}$$

The solutions x to (5.2) that also satisfy $x > (2k-1)AB/(1-\varepsilon)$ are seen, by substituting this bound into the right hand side of (5.2), to be such that

$$\sqrt{x} \;\leq\; \frac{A}{1-\varepsilon}\left[\left(1+\sqrt{\frac{B(1-\varepsilon)}{A(2k-1)}}\right)^{2k-1} - 1\right] + \frac{\|T\|_{2k}}{1-\varepsilon}. \tag{5.3}$$

Since, expanding the power, inequality (5.3) is satisfied by all $x \leq (2k-1)AB/(1-\varepsilon)$, it follows that all solutions of (5.2) satisfy (5.3), proving the first inequality. The second follows from the fact that, for $t \geq 0$ and $\gamma \geq 0$, $(1+t)^{\gamma} - 1 \leq \gamma t e^{\gamma t}$. □

Proof of Theorem 2.2. First, take $r = 2k$ with $k \in \mathbb{N}$. Then, using (2.1) with $f(w) = (w-\mu)^{2k-1}$, we have

$$\mathbb{E}[G\{f(D+W) - f(W)\}] \;=\; \mathbb{E}(W-\mu)^{2k} + \mathbb{E}\{R(W-\mu)^{2k-1}\}.$$

Invoking the assumption (2.2) and Hölder's inequality thus leads to

$$\begin{aligned}(1-\varepsilon)\,\mathbb{E}(W-\mu)^{2k} \;&\leq\; |\mathbb{E}[G\{f(D+W) - f(W)\}]| + \mathbb{E}\big\{T\big|W-\mu\big|^{2k-1}\big\}\\ &\leq\; |\mathbb{E}[G\{f(D+W) - f(W)\}]| + \sigma^{-1}\|T\|_{2k}\{\sigma\|W-\mu\|_{2k}^{2k-1}\}.\end{aligned}$$

Now, for this choice of f, using the fundamental theorem of calculus, we have

$$\begin{aligned}|f(D+W) - f(W)| \;&\leq\; (2k-1)|D|\max\{(W-\mu)^{2k-2}, (W'-\mu)^{2k-2}\}\\ &\leq\; (2k-1)|D|\{(W-\mu)^{2k-2} + (W'-\mu)^{2k-2}\}.\end{aligned}$$

Hence, by Hölder's inequality, and because $\mathbb{E}\{(W'-\mu)^{2k}\} \leq \mathbb{E}\{(W-\mu)^{2k}\}$, it follows that

$$\begin{aligned}(1-\varepsilon-\varepsilon')\|W-\mu\|_{2k}^{2k} \;&\leq\; (2k-1)\|G\|_{2k}\|D\|_{2k}\{\|W-\mu\|_{2k}^{2k-2} + \|W'-\mu\|_{2k}^{2k-2}\}\\ &\leq\; 2(2k-1)\|G\|_{2k}\|D\|_{2k}\|W-\mu\|_{2k}^{2k-2},\end{aligned}$$

and the theorem is proved for $r = 2k$.

For $r = 2k+1$, take $f(w) = (w-\mu)^{2k}\operatorname{sgn}(w-\mu)$, and observe that now

$$\begin{aligned}|f(D+W) - f(W)| \;&\leq\; 2k|D|\max\{|W-\mu|^{2k-1}, |W'-\mu|^{2k-1}\}\\ &\leq\; 2k|D|\{|W-\mu|^{2k-1} + |W'-\mu|^{2k-1}\};\end{aligned}$$

the remainder of the argument is the same. □

Proof of Theorem 2.3. Much as in the proof of Theorem 2.1, we begin by observing that

$$\begin{aligned}&\mathbb{E}\big\{(W-\mu)^{2k}\big\} + \mathbb{E}\big\{R(W-\mu)^{2k-1}\big\}\\ &=\; \sum_{j=1}^{2k-1}\binom{2k-1}{j}\mathbb{E}\big\{GD^{j}(W-\mu)^{2k-1-j}\big\}\\ &=\; (2k-1)\sigma^2\,\mathbb{E}\big\{(W-\mu)^{2k-2}\big\} + (2k-1)\,\mathbb{E}\big\{\big(\mathbb{E}(GD\,|\,W) - \sigma^2\big)(W-\mu)^{2k-2}\big\}\\ &\qquad + \sum_{j=2}^{2k-1}\binom{2k-1}{j}\mathbb{E}\big\{GD^{j}(W-\mu)^{2k-1-j}\big\}.\end{aligned}$$

Hence, using the extra assumption, we have

$$\begin{aligned}(1-\varepsilon_1-\varepsilon_2-(2k-1)\varepsilon_3)\,\mathbb{E}\big\{(W-\mu)^{2k}\big\} \\ \leq\ (2k-1)\sigma^2\,\mathbb{E}\big\{(W-\mu)^{2k-2}\big\}+(2k-1)\,\mathbb{E}\big\{T_2\{W-\mu\}^{2k-2}\big\} \\ +\sum_{j=2}^{2k-1}\binom{2k-1}{j}\mathbb{E}\big|GD^j(W-\mu)^{2k-1-j}\big|.\end{aligned} \tag{5.4}$$

Defining $x := \|W-\mu\|_{2k}^2 \geq \sigma^2$, and bounding the final sum as in the proof of Theorem 2.1, we deduce that

$$\begin{aligned}(1-\varepsilon_1-\varepsilon_2-(2k-1)\varepsilon_3)x^k\ \leq\ (2k-1)\sigma^2x^{k-1}+(2k-1)\|T_2\|_kx^{k-1} \\ +Ax^{k-1/2}\bigg\{\Big(1+\frac{B}{\sqrt{x}}\Big)^{2k-1}-1-\frac{(2k-1)B}{\sqrt{x}}\bigg\}.\end{aligned}$$

Using the fact that, for $a\geq 2$ and $y>0$, we have

$$(1+y)^a-1-ay\ \leq\ \tfrac{1}{2}a(a-1)y^2(1+y)^{a-2}\ \leq\ \tfrac{1}{2}a(a-1)y^2e^{ay}, \tag{5.5}$$

it follows that

$$\begin{aligned}\big(1-\varepsilon_1-\varepsilon_2-(2k-1)\varepsilon_3\big)x \\ \leq\ (2k-1)\bigg\{\sigma^2\bigg(1+\frac{(k-1)AB^2}{\sigma^2\sqrt{x}}\exp\Big\{\frac{(2k-1)B}{\sqrt{x}}\Big\}\bigg)+\|T_2\|_k\bigg\},\end{aligned}$$

from which, considering first the case $x\geq(2k-1)\sigma^2$, it follows that

$$x\ \leq\ \frac{(2k-1)\sigma^2}{1-\varepsilon_1-\varepsilon_2-(2k-1)\varepsilon_3}\bigg(1+\frac{(k-1)AB^2}{\sigma^3\sqrt{2k-1}}\exp\Big\{\frac{B\sqrt{2k-1}}{\sigma}\Big\}+\frac{\|T_2\|_k}{\sigma^2}\bigg),$$

as claimed. □

Proof of Theorem 2.4. We begin from (5.4). Defining $x_k := \|W-\mu\|_{2k}^2 \geq \sigma^2$, we now deduce that

$$(1-E)x_k^k\ \leq\ (2k-1)\sigma^2x_{k-1}^{k-1}+Ax_k^{k-1/2}\bigg\{\Big(1+\frac{B}{\sqrt{x_k}}\Big)^{2k-1}-1-\frac{(2k-1)B}{\sqrt{x_k}}\bigg\}. \tag{5.6}$$

Using the Taylor expansion argument (5.5), we have

$$Ax^{k-1/2}\bigg\{\Big(1+\frac{B}{\sqrt{x}}\Big)^{2k-1}-1-\frac{(2k-1)B}{\sqrt{x}}\bigg\}\ \leq\ x^kH_k(x),$$

where

$$H_k(x)\ :=\ x^{-3/2}(2k-1)(k-1)AB^2\exp\{(2k-1)B/\sqrt{x}\}.$$

Because $\sigma\geq B\sqrt{e(2k-1)}$, and from Lemma 2.5, we have

$$H_k(\sigma^2\|N\|_{2k}^2)\ \leq\ h_k.$$

It thus follows from (5.6) that either $\sqrt{x_k}\leq\sigma\|N\|_{2k}$ or

$$(1-E-h_k)x_k^k\ \leq\ (2k-1)\sigma^2x_{k-1}^{k-1}.$$

Using the same argument, and because h_k is increasing in k, we deduce that, for each $2\leq l\leq k$,

$$x_l^l\ \leq\ \max\bigg\{\frac{(2l-1)\sigma^2}{1-E-h_k}x_{l-1}^{l-1},(\sigma\|N\|_{2l})^{2l}\bigg\}.$$

Iterating this inequality, starting with $l=k$ and working downwards, and noting that $\|N\|_{2l}^{2l}=(2l-1)\|N\|_{2(l-1)}^{2(l-1)}$, we obtain

$$x_k^k\ \leq\ (1-E-h_k)^{-k}(\sigma\|N\|_{2k})^{2k},$$

proving the theorem. □

5.2 Proofs for concentration inequalities

Proof of Corollary 3.2. If $|D| \leq x_1$ and $|n^{-1}G| \leq x_2$ a.s. for all n, in order to exploit (3.2), we start by minimizing

$$\log\left\{t^{-2k}\left\{2knx_1x_2\right\}^k\right\} \;=\; -2k\log t + k\{\log n + \log(2k) + \log(x_1x_2)\}$$

in k. The calculus minimum is attained when

$$\log n + \log(2k) + \log(x_1x_2) - 2\log t + 1 \;=\; 0.$$

Thus, choosing $k := k_n(t) := \left\lfloor \frac{1}{2}\{t^2/(nx_1x_2e)\}\right\rfloor$, this gives

$$(2k)^k \;\leq\; \left\{\frac{t^2}{nx_1x_2e}\right\}^k,$$

and hence

$$t^{-2k}\left\{n(2k-1)x_1x_2\right\}^k \;\leq\; e^{-k} \;\leq\; e\exp\{-\tfrac{1}{2}t^2/(nx_1x_2e)\}.$$

Substituting this into (3.2) gives

$$\mathbb{P}[|W_n - \mu_n| > t] \;\leq\; e\exp\left\{-\frac{t^2}{2nx_1x_2e}\left(1 - \frac{2}{\sqrt{n}}\sqrt{\frac{(2k-1)x_1}{x_2}}\right)\right\}, \tag{5.7}$$

with k as above. Since

$$2k - 1 \;\leq\; 2k \;\leq\; \frac{t^2}{nx_1x_2e},$$

(5.7) is bounded as claimed. □

Proof of Corollary 3.3. Considering (3.2), we note that we can take

$$\alpha_{n,k} \;=\; \beta_{n,k} \;=\; \left\{\frac{2kc}{a}\left(b^{-1/a}\right)^{2k}\Gamma(2k/a)\right\}^{1/2k},$$

from Lemma 3.1. We then choose k to (approximately) minimize

$$-2k\log t + k\{\log n + \log(2k)\} + \log\left\{\frac{2kc}{a}\left(b^{-1/a}\right)^{2k}\Gamma(2k/a)\right\},$$

or, more simply, to (approximately) minimize, using a crude approximation of the gamma function along the lines of Stirling's formula,

$$-2k\log t + k\{\log n + \log(2k)\} - (2k/a)\log b + (2k/a)\log(2k/a) - 2k/a.$$

The calculus minimum is attained when

$$\log n + \log(2k) + 1 - 2\log t + (2/a)\{-\log b + \log(2k/a)\} \;=\; 0,$$

suggesting the choice, provided that $t^2 \geq ne2^{(a+2)/a}(ab)^{-2/a}$ so that $k \geq 1$, of

$$k \;=\; \left\lfloor \tfrac{1}{2}\left(\frac{t^2}{ne}\right)^{a/(a+2)}(ab)^{2/(a+2)}\right\rfloor, \tag{5.8}$$

for which

$$(2k)^{k(a+2)/a} \;\leq\; \left\{\frac{t^2}{ne}\right\}^k (ab)^{2k/a}. \tag{5.9}$$

Then, using the bound on the gamma function from [4, Corollary 1.2], the right hand side of (3.2) is at most

$$\begin{aligned} & t^{-2k} n^k (2k)^k c \left(b^{-1/a}\right)^{2k} \Gamma(2k/a+1) \exp\{n^{-1/2}(2k)^{3/2}\} \\ & \quad \le\ t^{-2k} n^k (2k)^k c \left(b^{-1/a}\right)^{2k} \sqrt{2\pi} \left(\frac{2k}{a}\right)^{2k/a} \left\{\frac{2k}{a}+\frac{1}{2}\right\}^{1/2} e^{-2k/a} \exp\{n^{-1/2}(2k)^{3/2}\} \\ & \quad \le\ c\sqrt{2\pi} \left\{\frac{2k}{a}+\frac{1}{2}\right\}^{1/2} e^{-k(a+2)/a} \exp\{n^{-1/2}(2k)^{3/2}\}, \end{aligned}$$

because of (5.9). The result follows after noting that k from (5.8) equals k_t from the statement of the corollary. □

Proof of Corollary 3.4. Using the upper and lower factorial bounds given by [23], we have

$$\|N\|_{2k}^{2k} \ =\ \frac{(2k)!}{k!2^k} \ \le\ \exp\{k\log k + (k+\tfrac{1}{2})\log 2 - k\},$$

and so

$$(2k)^{-k}\|N\|_{2k}^{2k} \ \le\ \exp\{\tfrac{1}{2}\log 2 - k\} \ =\ \sqrt{2}\, e^{-k}.$$

Let $k = \lfloor y^2/2 \rfloor$ and take $t = \sigma_n\sqrt{2k}$ in (3.1); then, using Theorem 2.4 and the previous display,

$$\mathbb{P}[|W_n-\mu_n| > y\sigma_n] \ \le\ \mathbb{P}[|W_n-\mu_n| > \sigma_n\sqrt{2k}] \ \le\ \frac{\sqrt{2}\, e^{-k}}{(1-E-h_k)^k} \ \le\ \frac{e\sqrt{2}\, e^{-y^2/2}}{(1-E-h_k)^k},$$

proving the corollary. □

5.3 Proofs for applications

Independent sums

Proof of Proposition 4.1. As in the proof of Theorem 2.1 (but with $R = 0$), we begin with

$$\mathbb{E}\big\{W^{2k}\big\} \ =\ \sum_{j=1}^{2k-1} \binom{2k-1}{j} (-1)^j\, \mathbb{E}\big\{nX_I^{j+1} W^{2k-1-j}\big\}. \tag{5.10}$$

For the first term in the sum, we have

$$(2k-1)\,\mathbb{E}\big\{nX_I^2 W^{2k-2}\big\} \ =\ (2k-1)\sum_{i=1}^{n} \mathbb{E}\big\{X_i^2 W^{2k-2}\big\}.$$

By the independence of X_i and $W_i := W - X_i$, we have

$$\mathbb{E}\big\{(X_i^2-\sigma_i^2)W^{2k-2}\big\} \ =\ \mathbb{E}\big\{(X_i^2-\sigma_i^2)\{W^{2k-2} - W_i^{2k-2}\}\big\},$$

and so, using the fundamental theorem of calculus, we have

$$\begin{aligned} & \big|\mathbb{E}\big\{(X_i^2-\sigma_i^2)W^{2k-2}\big\}\big| \\ & \qquad \le\ \mathbb{E}\big\{|X_i^2-\sigma_i^2||X_i|(2k-2)\{|W|^{2k-3} + |W_i|^{2k-3}\}\big\}. \end{aligned}$$

Now, Jensen's inequality implies

$$\mathbb{E}|W_i|^{2k} = \mathbb{E}\Big[\big|W_i + \mathbb{E}[X_i|W_i]\big|^{2k}\Big] \le \mathbb{E}\Big[\mathbb{E}\big[|W_i+X_i|^{2k}|W_i\big]\Big] = \mathbb{E}\big[|W|^{2k}\big],$$

and using this with Hölder's inequality yields

$$\mathbb{E}\big\{|X_i^3|\{|W|^{2k-3}+|W_i|^{2k-3}\}\big\} \;\le\; 2\|X_i\|_{2k}^3\|W\|_{2k}^{2k-3},$$

and $\sigma_i^2\,\mathbb{E}\big\{|X_i|\{|W|^{2k-3}+|W_i|^{2k-3}\}\big\}$ is bounded by the same quantity. Hence

$$\begin{aligned}(2k-1)\big|\mathbb{E}\big\{nX_I^2W^{2k-2}\big\}-\sigma^2\,\mathbb{E}\big\{W^{2k-2}\big\}\big| \\ \le\; 8(k-1)(2k-1)\Big\{\sum_{i=1}^n\|X_i\|_{2k}^3\Big\}\|W\|_{2k}^{2k-3} \\ \le\; 8n(k-1)(2k-1)\|X_I\|_{2k}^3\|W\|_{2k}^{2k-3}.\end{aligned}$$

The remaining terms in the sum in (5.10) are equal to

$$nX_I\big[(W-X_I)^{2k-1}-W^{2k-1}+(2k-1)X_IW^{2k-2}\big],$$

which can be bounded in absolute value, using Taylor's theorem with integral remainder, by

$$\begin{aligned}(2k-1)(k-1)\,\mathbb{E}\big\{n|X_I|^3\{|W|^{2k-3}+|W'|^{2k-3}\}\big\} \\ \le\; 2n(k-1)(2k-1)\|X_I\|_{2k}^3\|W\|_{2k}^{2k-3}.\end{aligned}$$

We have thus shown that

$$\|W\|_{2k}^{2k} \;\le\; n(2k-1)\sigma^2\|W\|_{2k-2}^{2k-2}+\|W\|_{2k}^{2k}\bigg(10n^{-1/2}(k-1)(2k-1)\rho_k^3\frac{\sigma^3}{\|W\|_{2k}^3}\bigg).$$

So if $\|W\|_{2k}\ge\sigma\|N\|_{2k}$, then

$$10n^{-1/2}(k-1)(2k-1)\rho_k^3\frac{\sigma^3}{\|W\|_{2k}^3} \;\le\; 5\sqrt{2e^3}\rho_k^3\sqrt{\frac{k-1}{n}} \;=\; h_k'.$$

Arguing as in the proof of Theorem 2.4 completes the proof. □

Sums of local statistics of sparse Erdős–Rényi random graphs

Proof of Theorem 4.2. To obtain moment bounds for W, we define a Stein coupling. For each $j=1,\dots,n$, set

$$\mathbb{I}_r^{(j)} \;:=\; \big\{\{i,l\}\subset[n]\colon\,\{i,l\}\cap\mathcal{N}_r(j)\neq\emptyset\big\}.$$

Let $(E'_{il},\{i,l\}\subset[n])$ be independent indicators, independent also of $\mathcal{G}_n$, each with $\mathbb{E}E'_{il}=p$. Given $\mathcal{G}_n$, define the random graph $\mathcal{G}_n^{(j)}$ by replacing all the edge indicators $(E_{il},\,\{i,l\}\in\mathbb{I}_r^{(j)})$, between pairs of vertices with at least one vertex in the r-neighbourhood of j, with $(E'_{il},\,\{i,l\}\in\mathbb{I}_r^{(j)})$, leaving the edges between other pairs of vertices the same as those of $\mathcal{G}_n$. Define

$$X_i^{(j)} \;:=\; U\big(\mathcal{N}_r(i,\mathcal{G}_n^{(j)})\big),$$

and $W^{(j)}:=\sum_{i=1}^n X_i^{(j)}$. Finally, let J be uniform on the set $\{1,\dots,n\}$, independent of the random objects above, and define $W':=W^{(J)}$ and $G=-n(X_J-\mathbb{E}X_J)$. After noting that X_j is independent of $W^{(j)}$, it is easy to see that (W,W',G) is an exact Stein coupling. Moreover, $\mathcal{L}(W')=\mathcal{L}(W)$, so that the central moments of W and W' are equal, and hence both of Theorems 2.1 and 2.2 apply, with $\varepsilon=\varepsilon'=0$; we use Theorem 2.2.

To apply it to bound the q-th central moment, we need to bound

$$\|G\|_q^q \;=\; n^{q-1}\sum_{i=1}^n\mathbb{E}|X_i-\mathbb{E}X_i|^q, \tag{5.11}$$

and

$$\|D\|_q^q \;=\; \frac{1}{n}\sum_{j=1}^n \mathbb{E}\big|W^{(j)} - W\big|^q \\ \;=\; \frac{1}{n}\sum_{j=1}^n \mathbb{E}\Big|\sum_{i=1}^n (X_i^{(j)} - X_i)\Big|^q \tag{5.12}$$

To bound (5.11), define $N_r(i) := |V(\mathcal{N}_r(i,\mathcal{G}_n))|$ to be the number of vertices in the r-neighbourhood of i in $\mathcal{G}_n$, and note that, due to (4.7),

$$|X_i| \;\le\; cN_r(i)^\beta,$$

and in particular, from Lemma 5.3, for

$$A(x,\ell) \;:=\; \pi e^{e-2} \times \begin{cases} \ell/\log\big((e-1)\big), & \ell > x, \\ x, & \ell \le x, \end{cases}$$

we have

$$\|X_i\|_\ell \;\le\; c\|N_r(i)^\beta\|_\ell \;=\; c\|N_r(i)\|_{\beta\ell}^\beta \;\le\; 2^\beta c A(\lambda,\ell\beta)^{r\beta}. \tag{5.13}$$

Using (5.13) and Minkowski's inequality, this gives

$$\mathbb{E}\big|X_i - \mathbb{E}X_i\big|^q \;=\; \|X_i - \mathbb{E}X_i\|_q^q \;\le\; \big(\|X_i\|_q + \|X_i\|_1\big)^q \;\le\; (2^{1+\beta}c)^q\, A(\lambda,q\beta)^{r\beta q}.$$

Thus, from (5.11), we can bound

$$\|G\|_q \;\le\; 2^{1+\beta} nc (C_A \Lambda_q^{(1)})^{r\beta}, \tag{5.14}$$

where $\Lambda_s^{(1)} := \max\{\lambda, s\beta\}$ and $C_A = \pi e^{e-2}/\log(e-1)$.

Moving to (5.12), we note that $z_j := \mathbb{E}\big|\sum_{i=1}^n (X_i^{(j)} - X_i)\big|^q$ is the same for all j, so that $\|D\|_q^q$ in (5.12) can be bounded by z_1. Define the indicators

$$I_i \;:=\; I[E'_{il} \;=\; 1 \text{ for some } l \in V(\mathcal{N}_r(1,\mathcal{G}_n))],$$

and then the subsets of vertices

$$\mathbb{I}_-^{(1)} \;:=\; V\big(\mathcal{N}_{r+1}(1,\mathcal{G}_n)\big); \qquad \mathbb{I}_+^{(1)} \;:=\; \{i \notin V(\mathcal{N}_r(1,\mathcal{G}_n))\colon\, I_i = 1\}.$$

The difference $X_j^{(1)} - X_j$ can be written as $(X_j^{(1)} - \widehat{X}_j^{(1)}) + (\widehat{X}_j^{(1)} - X_j)$, where $\widehat{X}_j^{(1)} := U\big(\mathcal{N}_r(j,\mathcal{G}_{n,r}^{(1)})\big)$, and $\mathcal{G}_{n,r}^{(1)}$ is defined to be the subgraph of $\mathcal{G}_n$ consisting of the edges between the vertices $\{i \notin N_r(1,\mathcal{G}_n)\}$; in particular, vertices in $N_r(1,\mathcal{G}_n)$ are isolated in $\mathcal{G}_{n,r}^{(1)}$. Then $\widehat{X}_j^{(1)} - X_j$ can only be non-zero if $j \in \mathcal{N}_{r-1}(i,\mathcal{G}_n)$ for some $i \in \mathbb{I}_-^{(1)}$, and $X_j^{(1)} - \widehat{X}_j^{(1)}$ can only be non-zero if $j \in \mathcal{N}_{r-1}(i,\mathcal{G}_n)$ for some $i \in V\big(\mathcal{N}_r(1,\mathcal{G}_n)\big) \cup \mathbb{I}_+^{(1)}$. In the former case,

$$|\widehat{X}_j^{(1)} - X_j| \;\le\; 2cN_r(j,\mathcal{G}_n)^\beta \;\le\; 2cN_{3r}(1,\mathcal{G}_n)^\beta,$$

and there are at most $N_{2r}(1,\mathcal{G}_n)$ such indices; hence

$$\sum_{j=1}^n |\widehat{X}_j^{(1)} - X_j| \;\le\; 2cN_{2r}(1,\mathcal{G}_n)\{N_{3r}(1,\mathcal{G}_n)\}^\beta. \tag{5.15}$$

To bound $|X_j^{(1)} - \widehat{X}_j^{(1)}|$, define

$$\mathcal{N}_{lr}^*(1) \;:=\; \mathcal{N}_r(1,\mathcal{G}_n) \cup \widehat{\mathcal{N}}_{lr}(1),$$

where

$$\widehat{\mathcal{N}}_{lr}(1) \;:=\; \bigcup_{i\in\mathbb{I}_+^{(1)}} \mathcal{N}_{(l-1)r-1}(i,\mathcal{G}_{n,r}^{(1)}), \qquad l=2,3;$$

set $\widehat{N}_{lr}(1) := |V(\widehat{\mathcal{N}}_{lr}(1))|$, and note that $N^*_{lr}(1) := |V(\mathcal{N}^*_{lr}(1))| \le N_r(1) + \widehat{N}_{lr}(1)$. Then there are at most $N^*_{2r}(1)$ indices j at which $|X_j^{(1)} - \widehat{X}_j^{(1)}|$ can be non-zero, and, for such j, the difference cannot exceed

$$c\big|V\big(\mathcal{N}_r(j,G_n^{(1)})\big)\big|^\beta + c\{N_r(j)\}^\beta \;\le\; 2c\{N^*_{3r}(1)\}^\beta.$$

Hence

$$\sum_{j=1}^n |X_j^{(1)} - \widehat{X}_j^{(1)}| \;\le\; 2cN^*_{2r}(1)\{N^*_{3r}(1)\}^\beta. \tag{5.16}$$

Now, by Hölder's inequality,

$$\begin{aligned}\|N_{2r}(1)\{N_{3r}(1)\}^\beta\|_q \;&\le\; \|N_{2r}(1)\|_{(1+\beta)q}\,\|\{N_{3r}(1)\}^\beta\|_{(1+\beta)q/\beta}\\ &=\; \|N_{2r}(1)\|_{(1+\beta)q}\,\|N_{3r}(1)\|^\beta_{(1+\beta)q},\end{aligned}$$

which can be bounded using Lemma 5.3; this gives

$$\|N_{2r}(1)\{N_{3r}(1)\}^\beta\|_q \;\le\; 2^{1+\beta}A(\lambda,(1+\beta)q)^{2r+3r\beta} \;\le\; 2^{1+\beta}\{C_A\Lambda_q^{(2)}\}^{2r+3r\beta}, \tag{5.17}$$

where $\Lambda_s^{(2)} := \max\{\lambda, s(1+\beta)\}$. Similarly, we have

$$\|N^*_{2r}(1)\{N^*_{3r}(1)\}^\beta\|_q \;\le\; \|N^*_{2r}(1)\|_{(1+\beta)q}\,\|N^*_{3r}(1)\|^\beta_{(1+\beta)q}, \tag{5.18}$$

where $N^*_s(1) \le N_r(1) + \widehat{N}_s(1)$. The norms of $N_r(1)$ are bounded once more by Lemma 5.3. For $\widehat{N}_{lr}(1)$, $l=2,3$, we set $N^{(1)}_{(l-1)r-1}(i) := |V(\mathcal{N}_{(l-1)r-1}(i,\mathcal{G}^{(1)}_{n,r}))|$ and observe that, conditional on $\mathcal{N}_r(1,\mathcal{G}_n)$, the indicators $\{I_i,\, i\notin\mathcal{N}_r(1,\mathcal{G}_n)\}$ are independent of the quantities $\{N^{(1)}_{(l-1)r-1}(i),\, i\notin\mathcal{N}_r(1,\mathcal{G}_n)\}$; the I_i all have expectation less than $pN_r(1)$, and, for $i\notin\mathcal{N}_r(1,\mathcal{G}_n)$, each $N^{(1)}_{(l-1)r-1}(i)$ has distribution dominated by the *unconditional* distribution of $N_{(l-1)r-1}(i)$, because $\mathcal{N}_{(l-1)r-1}(i,\mathcal{G}^{(1)}_{n,r})$ is constructed just as $\mathcal{N}_{(l-1)r-1}(i,\mathcal{G}_n)$, but from a subset of the vertices $[n]$. Applying Lemma 5.1, with $\mathcal{E} := \mathbb{I}_+^{(1)}$ and $Y_i := N^{(1)}_{(l-1)r-1}(i)$, it follows that

$$\mathbb{E}\Big(\widehat{N}_{lr}(1)^s \,|\, \widehat{N}_r(1)\Big) \;\le\; \mathbb{E}\big\{\widehat{B}^s \,|\, N_r(1)\big\}\;\mathbb{E}\Big\{\big(N_{(l-1)r-1}(i)\big)^s\Big\},$$

where $\widehat{B}\sim \mathrm{Bi}(n,pN_r(1))$. Bounding the expectations by Lemmas 5.2 and 5.3, this in turn gives

$$\begin{aligned}\mathbb{E}\Big(\widehat{N}_{lr}(1)^s \,|\, N_r(1)\Big) \;&\le\; A(\lambda N_r(1),s)^s\{2A(\lambda,s)^{(l-1)r-1}\}^s\\ &\le\; \{C_A(\lambda N_r(1)+s)\}^s\{2A(\lambda,s)^{(l-1)r-1}\}^s.\end{aligned}$$

Invoking Lemma 5.3 once again, it thus follows that

$$\begin{aligned}\|\widehat{N}_{lr}(1)\|_s \;&\le\; C_A\big\{\lambda\|N_r(1)\|_s + s\big\}\,\{2A(\lambda,s)^{(l-1)r-1}\}\\ &\le\; C_A\big\{2\lambda A(\lambda,s)^r + s\big\}\,\{2A(\lambda,s)^{(l-1)r-1}\}\\ &\le\; 8\{C_A(\lambda\vee s)\}^{lr}.\end{aligned}$$

Altogether we find that, for $s \geq 1$ and $l = 2, 3$,

$$\|N^*_{lr}(1)\|_s \leq \|N_r(1)\|_s + \|\widehat{N}_{lr}(1)\|_s \leq 2\big\{C_A(\lambda \vee s)\big\}^r + 8\{C_A(\lambda \vee s)\}^{lr} \leq 10\{C_A(\lambda \vee s)\}^{lr}.$$

Hence, from (5.18),

$$\|N^*_{2r}(1)\{N^*_{3r}(1)\}^\beta\|_q \;\leq\; 10^{1+\beta}\{C_A\Lambda^{(2)}_q\}^{2r+3r\beta}. \tag{5.19}$$

Combining (5.17) and (5.19) with (5.12), (5.15) and (5.16), we see that

$$\|D\|_q \;\leq\; (10^{1+\beta} + 2^{1+\beta})c(C_A\Lambda^{(2)}_q)^{2r+3r\beta}. \tag{5.20}$$

Now, applying Theorem 2.2 with the bounds (5.14) and (5.20), the theorem follows. □

Lemma 5.1. *Let $\mathcal{I}$ be a finite index set, $\mathcal{E}$ be a random (possibly empty) subset of $\mathcal{I}$, and define $E := |\mathcal{E}|$. Let $\{Y_i\}_{i\in\mathcal{I}}$ be a collection of random variables independent of $\mathcal{E}$ and, for $\ell \in \mathbb{N}$, let $\max_{i\in\mathcal{I}} \|Y_i\|_\ell \leq y$. Then*

$$\Big\|\sum_{i\in\mathcal{E}} Y_j\Big\|_\ell \;\leq\; y\|E\|_\ell.$$

Proof. Using Minkowski's inequality in the third line, we have

$$\begin{aligned}
\Big\|\sum_{i\in\mathcal{E}} Y_i\Big\|^\ell_\ell \;&=\; \mathbb{E}\Big[\Big(\sum_{i\in\mathcal{E}} Y_i\Big)^\ell\Big] \\
&=\; \sum_{\mathcal{S}\subseteq\mathcal{I}} \mathbb{E}\Big[\Big(\sum_{i\in\mathcal{S}} Y_i\Big)^\ell \big|\mathcal{E} = \mathcal{S}\Big]\,\mathbb{P}(\mathcal{E} = \mathcal{S}) \\
&\leq\; \sum_{\mathcal{S}\subseteq\mathcal{I}} \Big(\sum_{i\in\mathcal{S}} \|Y_i\|_\ell\Big)^\ell\,\mathbb{P}(\mathcal{E} = \mathcal{S}) \\
&\leq\; y^\ell \sum_{\mathcal{S}\subseteq\mathcal{I}}^{\infty} |\mathcal{S}|^\ell\,\mathbb{P}(\mathcal{E} = \mathcal{S}) \;=\; y^\ell\|E\|^\ell_\ell.
\end{aligned}$$

□

Lemma 5.2. *Let $n \in \mathbb{N}$, $0 \leq p \leq 1$, $Y \sim \mathrm{Bi}(n, p)$, and $\ell \in \mathbb{N}$. Then $\|Y\|_\ell \leq A(np, \ell)$, where*

$$A(x, \ell) \;:=\; \pi e^{e-2} \times \begin{cases} \ell/\log\big((e-1)\big), & \ell > x, \\ x, & \ell \leq x. \end{cases}$$

In particular, $A(x, l) \leq C_A(x \vee l) \leq C_A(x + l)$, where $C_A := \pi e^{e-2}/\log(e-1)$.

Proof. For any $\theta > 0$, we have, using Markov's inequality,

$$\begin{aligned}
\mathbb{E}[Y^\ell] \;&=\; \ell \int_0^\infty t^{\ell-1}\,\mathbb{P}(Y \geq t)dt \\
&\leq\; \ell \int_0^\infty t^{\ell-1}\,\mathbb{E}[e^{\theta Y}]e^{-\theta t}dt \\
&=\; \Gamma(\ell+1)\theta^{-\ell}\big(1 - p + pe^\theta\big)^n.
\end{aligned} \tag{5.21}$$

If $\ell > np$, we choose $\theta = \log\big((e-1)\ell/(np)\big)$, which yields

$$\|Y\|_\ell \;\leq\; \Gamma(\ell+1)^{1/\ell}\frac{\big(1 - p + \frac{(e-1)\ell}{n}\big)^{n/\ell}}{\log\big((e-1)\ell/(np)\big)} \;\leq\; \Gamma(\ell+1)^{1/\ell}\frac{e^{e-1}}{\log\big((e-1)\ell/(np)\big)}.$$

Now using the explicit bound on the Gamma function given in [4, Corollary 1.2]:

$$\Gamma(\ell+1) \;\leq\; \sqrt{\pi(2\ell+1)}\ell^{\ell}e^{-\ell}, \tag{5.22}$$

we find

$$\|Y\|_\ell \;\leq\; \frac{e^{e-2}\ell}{\log\bigl(\frac{(e-1)\ell}{np}\bigr)}\bigl(\pi(2\ell+1)\bigr)^{1/(2\ell)} \;\leq\; \frac{e^{e-2}\pi\ell}{\log(e-1)},$$

where the second inequality follows from elementary considerations, noting $\ell > np$.

If $\ell \leq np$, set $\theta = \ell/(np)$ in (5.21) and use (5.22) to find

$$\|Y\|_\ell \;\leq\; np\frac{\Gamma(\ell+1)^{1/\ell}}{\ell}\bigl(1-p+pe^{\ell/(np)}\bigr)^{n/\ell} \;\leq\; np\bigl(\pi(2\ell+1)\bigr)^{1/(2\ell)}e^{-1}e^{\frac{np}{\ell}\left(e^{\ell/(np)}-1\right)}.$$

Recalling that $\ell \leq np$, the final exponential is maximized at $\ell = np$, which leads to

$$\|Y\|_\ell \;\leq\; np\bigl(\pi(2\ell+1)\bigr)^{1/(2\ell)}e^{e-2} \;\leq\; e^{e-2}\pi(np). \qquad \square$$

Lemma 5.3. *Fix $n, r \in \mathbb{N}$ and $p \in [0,1]$, and let N_r be the number of vertices in the r-neighbourhood of a given vertex in an Erdős–Rényi random graph on n vertices with edge probability parameter p. Then*

$$\|N_r\|_\ell \;\leq\; \frac{A(np,\ell)^{r+1}-1}{A(np,\ell)-1} \;\leq\; 2A(np,\ell)^r,$$

where $A(x,\ell)$ is defined in Lemma 5.2.

Proof. Let $(Z_0, Z_1, Z_2, \ldots)$ be the sequence of generation sizes for a Galton-Watson branching process with offspring distribution $\mathrm{Bi}(n,p)$, started from a single individual; $Z_0 = 1$. Then, using the usual exploration process process coupling found, for example, in [25, Chapter 4], $\sum_{s=0}^{r} Z_s$ stochastically dominates N_r, so that $\|N_r\|_\ell \leq \sum_{s=0}^{r}\|Z_s\|_\ell$, for any $r, \ell \geq 0$. To bound $\|Z_s\|_\ell$, we use the random sum representation of a branching process and repeatedly apply Lemma 5.1 to find

$$\|Z_s\|_\ell \;\leq\; \|B\|_\ell^s,$$

where $B \sim \mathrm{Bi}(n,p)$. The result now easily follows from Lemma 5.2, and because $A(x,l) \geq 2$ for all $x > 0$ and $l \in \mathbb{N}$. $\square$

References

[1] R. Arratia and P. Baxendale, 2015. Bounded size bias coupling: a Gamma function bound, and universal Dickman-function behavior. *Probab. Theory Related Fields*, 162(3-4):411–429. MR-3383333

[2] A. D. Barbour, L. Holst, and S. Janson, 1992. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. Oxford University Press, Oxford. MR-1163825

[3] A. D. Barbour, A. Röllin, and N. Ross, 2019. Error bounds in local limit theorems using Stein's method. *Bernoulli*, 25(2):1076–1104. MR-3920366

[4] N. Batir, 2008. Inequalities for the gamma function. *Arch. Math. (Basel)*, 91(6):554–563. MR-2465874

[5] S. Boucheron, G. Lugosi, and P. Massart, 2003. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614. MR-1989444

[6] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart, 2005. Moment inequalities for functions of independent random variables. *Ann. Probab.*, 33(2):514–560. MR-2123200

[7] S. Boucheron, G. Lugosi, and P. Massart, 2013. *Concentration inequalities*. Oxford University Press, Oxford. A nonasymptotic theory of independence, With a foreword by Michel Ledoux. MR-3185193

[8] S. Chatterjee, 2007. Stein's method for concentration inequalities. *Probab. Theory Related Fields*, 138(1-2):305–321. MR-2288072

[9] S. Chatterjee, 2014. A short survey of Stein's method. In S. Y. Jang, Y. R. Kim, D.-W. Lee, and I. Yie, editors, *Proceedings of the International Congress of Mathematicians, Seoul 2014, Volume IV, Invited Lectures*, pages 1–24, Seoul, Korea. KYUNG MOON SA Co. Ltd. MR-3727600

[10] S. Chatterjee and P. S. Dey, 2010. Applications of Stein's method for concentration inequalities. *Ann. Probab.*, 38(6):2443–2485. MR-2683635

[11] L. H. Y. Chen and A. Röllin, 2010. Stein couplings for normal approximation. Preprint arXiv:1003.6039v2.

[12] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao, 2011. *Normal approximation by Stein's method*. Probability and its Applications (New York). Springer, Heidelberg. MR-2732624

[13] L. H. Y. Chen, X. Fang, and Q.-M. Shao, 2013. From Stein identities to moderate deviations. *Ann. Probab.*, 41(1):262–293. MR-3059199

[14] N. Cook, L. Goldstein, and T. Johnson, 2018. Size biased couplings and the spectral gap for random regular graphs. *Ann. Probab.*, 46(1):72–125. MR-3758727

[15] S. Ghosh and L. Goldstein, 2011. Concentration of measures via size-biased couplings. *Probab. Theory Related Fields*, 149(1-2):271–278. MR-2773032

[16] S. Ghosh and L. Goldstein, 2011. Applications of size biased couplings for concentration of measures. *Electron. Commun. Probab.*, 16:70–83. MR-2763529

[17] S. Ghosh, L. Goldstein, and M. Raič, 2011. Concentration of measure for the number of isolated vertices in the Erdős–Rényi random graph by size bias couplings. *Statist. Probab. Lett.*, 81(11):1565–1570. MR-2832913

[18] S. Janson, 2004. Large deviations for sums of partly dependent random variables. *Random Structures Algorithms*, 24(3):234–248. MR-2068873

[19] L. Mackey, M. I. Jordan, R. Y. Chen, B. Farrell, and J. A. Tropp, 2014. Matrix concentration inequalities via the method of exchangeable pairs. *Ann. Probab.*, 42(3):906–945. MR-3189061

[20] P. Massart, 2007. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. MR-2319879

[21] C. McDiarmid, 1998. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin. MR-1678578

[22] D. Paulin, L. Mackey, and J. A. Tropp, 2016. Efron-Stein inequalities for random matrices. *Ann. Probab.*, 44(5):3431–3473. MR-3551202

[23] H. Robbins, 1955. A remark on Stirling's formula. *Amer. Math. Monthly*, 62:26–29. MR-0069328

[24] A. Röllin, 2007. Translated Poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.*, 17(5-6):1596–1614. MR-2358635

[25] R. van der Hofstad, 2017. *Random graphs and complex networks. Vol. 1*. Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge. MR-3617364

Acknowledgments. The authors were partially supported by the Australian Research Council Discovery Grant DP150101459 and the ARC Centre of Excellence for Mathematical and Statistical Frontiers, CE140100049. We thank the referee for their helpful comments, and Larry Goldstein and Peter Eichelsbacher for pointing out some errors and omissions in a previous draft.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe `http://www.lockss.org/`
[2]EJMS: Electronic Journal Management System `http://www.vtex.lt/en/ejms.html`
[3]IMS: Institute of Mathematical Statistics `http://www.imstat.org/`
[4]BS: Bernoulli Society `http://www.bernoulli-society.org/`
[5]Project Euclid: `https://projecteuclid.org/`
[6]IMS Open Access Fund: `http://www.imstat.org/publications/open.htm`