Exponential Approximation by Stein's Method and Spectral Graph Theory

Running head: Exponential Approximation by Stein's Method

Version of 8/16/08

By Sourav Chatterjee, Jason Fulman, and Adrian Röllin

Abstract: General Berry-Esséen bounds are developed for the exponential distribution using Stein's method. As an application, a sharp error term is obtained for Hora's result that the spectrum of the Bernoulli-Laplace Markov chain has an exponential limit. This is the first use of Stein's method to study the spectrum of a graph with a non-normal limit.

2000 Mathematics Subject Classification: 60C05, 60F05.

Key words and phrases: Stein's method, spectral graph theory, Markov chain, exponential distribution.

1. Introduction

This paper develops general Berry-Esséen bounds for the exponential distribution using Stein's method. Two of our main results are given by the following statements. We let $\mathbb{I}[A]$ denote the indicator function of an event A.

Theorem 1.1. *Assume that W and W' are non-negative random variables on the same probability space such that $\mathcal{L}(W') = \mathcal{L}(W)$. Then, if $Z \sim \mathrm{Exp}(1)$, we have for any $t > 0$ and any constant $\lambda > 0$*

$$\begin{aligned}|\mathbb{P}[W \le t] - \mathbb{P}[Z \le t]| \;\le\; & \mathbb{E}\left|(\lambda^{-1}\mathbb{E}(D|W) + 1)\mathbb{I}[W > 0]\right| + \mathbb{E}\left|\tfrac{1}{2\lambda}\mathbb{E}(D^2|W) - 1\right| \\ & + \tfrac{1}{6\lambda}\mathbb{E}\,|D|^3 + \tfrac{1}{2\lambda}\mathbb{E}\left(D^2\mathbb{I}[|W-t| \le |D|]\right),\end{aligned}$$

where $D := W' - W$.

Theorem 1.2. *Assume that W and W' are non-negative random variables on the same probability space such that $\mathcal{L}(W') = \mathcal{L}(W)$ and*

$$\mathbb{E}(D|W) = -\lambda(W-1),$$

where $\lambda > 0$ is a fixed constant. Then if $Z \sim \mathrm{Exp}(1)$, we have for any $t > 0$,

$$\begin{aligned}|\mathbb{P}[W \le t] - \mathbb{P}[Z \le t]| \le \frac{\mathbb{E}\left|2\lambda W - \mathbb{E}(D^2|W)\right|}{2\lambda t} + \frac{\mathbb{E}|D|^3 \max\{t^{-1}, 2t^{-2}\}}{6\lambda} \\ + \frac{\mathbb{E}\left\{D^2\mathbb{I}\left[|W-t| \le |D|\right]\right\}}{\lambda t}.\end{aligned}$$

where $D := W' - W$.

The use of a pair (W, W') is similar to the exchangeable pairs approach of Stein for normal approximation [St1], but in the spirit of [Ro], throughout this paper we require only the weaker assumption that W and W' have the same law. It can be challenging to obtain good bounds on the error terms in Theorems 1.1 and 1.2, and we also develop a number of tools for doing that.

Before continuing, we mention that this is not the first paper to study exponential approximation by Stein's method. Indeed, earlier works, in the more general context of chi-squared approximation, include Mann [Mn], Luk [Lu], and Reinert [Re] (which also includes a discussion of unpublished work of Pickett). The paper [Mn] uses exchangeable pairs, whereas [Lu] and [Re] use the generator approach to Stein's method. However all of these papers focus on approximating expectations of smooth functions of W, rather than indicator functions of intervals, and so do not give Berry-Esséen theorems. Moreover, the examples in [Lu] and [Re] are about sums of independent random variables, whereas our example involves dependence.

Our main example is the spectrum of the Bernoulli-Laplace Markov chain. This Markov chain was suggested as a model of diffusion and has the following description. Let n be even. There are two urns, the first containing $\frac{n}{2}$ white balls, and the second containing $\frac{n}{2}$ black balls. At each stage, a ball is picked at random from each urn and the two are switched. Diaconis and Shahshahani [DS] proved that $\frac{n}{8}\log(n) + \frac{cn}{2}$ steps suffice for this process to reach equilibrium, in the sense that the total variation distance to the stationary distribution is at most ae^{-dc} for positive universal constants a and d. In order to prove this, they used the fact that the spectrum of the Markov chain consists of the numbers $1 - \frac{i(n-i+1)}{(n/2)^2}$ occurring with multiplicity $\binom{n}{i} - \binom{n}{i-1}$ for $1 \leq i \leq \frac{n}{2}$ and multiplicity 1 if $i = 0$. Hora proved the following result, which shows that the spectrum of the Bernoulli-Laplace chain has an exponential limit.

Theorem 1.3. *(*[Ho1]*) Consider the uniform measure on the set of the $\binom{n}{\frac{n}{2}}$ eigenvalues of the Bernoulli-Laplace Markov chain. Let τ be a random eigenvalue chosen from this measure. Then as $n \to \infty$, the random variable $W := \frac{n}{2}\tau + 1$ converges in distribution to an exponential random variable with mean* 1.

As an application of our general Berry-Esséen bound, the following result will be proved.

Theorem 1.4. *Let $Z \sim \mathrm{Exp}(1)$, and let W be as in Theorem 1.3. Then*

$$|\mathbb{P}\{W \leq t\} - \mathbb{P}\{Z \leq t\}| \leq \frac{C}{\sqrt{n}},$$

for all t, where C is a universal constant. Moreover this rate is sharp in the sense that there is a sequence of n's tending to infinity, and corresponding t_n's such that

$$|\mathbb{P}(W \leq t_n) - \mathbb{P}(Z \leq t_n)| = \frac{2e^{-2}}{\sqrt{n}} + O(1/n).$$

Note that the Bernoulli-Laplace Markov chain is equivalent to random walk on the Johnson graph $J(n,k)$ where $k = \frac{n}{2}$. The vertices consist of all size k subsets of $\{1, \cdots, n\}$, and two subsets are connected by an edge if they differ in exactly one element. From a given vertex, random walk on the Johnson graph picks a neighbor uniformly at random, and moves there.

One reason why our method for proving Theorem 1.4 is of interest (despite the existence of a more elementary argument for a weaker version of Theorem 1.4 sketched at the end of Section 4) is that Theorem 1.4 is in fact a small piece of a much larger program. To explain, limit theorems for graph spectra (especially Cayley graphs and finite symmetric spaces) have been studied by many authors and from various perspectives; some references are [Ho1], [Ho2], [Ke], [F1], [F2], [F3], [F4], [Sn], [ShSu], [T1], and [T2]. In particular, the references [T1] and [T2] describe some challenging conjectures where the limit distribution is the semicircle law and relate them to deep work in number theory. With the long-term goal of making progress on these conjectures, the paper [F4] gave some general constructions for applying Stein's method to study graph spectra, and worked out examples in the case of normal approximation. The current paper works out an exponential example, and is excellent evidence that these constructions will prove useful in other settings where the spectrum has a non-normal limit. We also emphasize that while there are papers such as [GT] which obtain non-normal limit theorems with an error term in spectral problems, they study the spectrum of random objects, whereas our work, and the conjectures of [T1], [T2], all pertain to the spectrum of a sequence of fixed, non-random graphs.

We also mention that an additional reason for studying the spectrum of the Bernoulli-Laplace chain is that it is closely related to the spectrum of the random transposition walk (and so with representation theory of the symmetric group). Indeed, from [Sc], the eigenvalues of the Bernoulli-Laplace chain can be expressed as $\frac{2(n-1)}{n}\mu - \frac{(n-2)}{n}$ as μ ranges over a subset of eigenvalues of the random transposition walk. This relationship is not surprising given that the Bernoulli-Laplace chain transposes balls from different urns at each step. But together with the large body of work on Kerov's central limit theorem for the spectrum of the random transposition walk ([Ke], [F2], [F3], [F4], [Sn], [IO], [Ho2]), it does make the problems studied in the current paper very natural. As a final justification for the current paper, we believe that the example in it will serve as a useful testing ground for other researchers in Stein's method (certainly it helped us in developing our Berry-Esséen theorems).

The organization of this paper is as follows. Section 2 proves our first general Berry-Esséen bound for the exponential law, namely Theorem 1.1 above, and develops tools for analyzing the error terms which appear in it. Section 3 proves our second general Berry-Esséen bound for the exponential law, namely Theorem 1.2 above, and develops tools paralleling those in Section 2 for analyzing the error terms. Section 4 treats our main example (spectrum of the Bernoulli-Laplace chain), proving Theorem 1.4. An interesting feature of the proof is that it uses theory from both of Sections 2 and 3, to treat the cases of small and large t respectively. Finally, Appendix A gives an algebraic approach to the exchangeable pair and moment computations in Section 4, linking it with the constructions of [F4]. This is not essential to the proofs of any of the results in the main body of the paper, but does motivate the exchangeable pairs used in the paper, which could be difficult to guess.

2. Berry-Esséen Bound for the Exponential Law: Version 1

A main purpose of this section is to prove Theorem 1.1 from the introduction, and to develop tools for analyzing the error terms which arise in it. To begin we make some remarks concerning the statement of Theorem 1.1.

Remarks:

(1) In our main example (see Section 4), the relation $\mathbb{E}(D|W) = -\lambda$ is satisfied for all $W > 0$. Hence the first error term in Theorem 1.1 will vanish. In the spirit of [RR], one could also have that $\mathbb{E}(D|W) = -\lambda + R$ for some non-trivial random variable R.

(2) Although W is allowed to attain the value 0 (and does, in our main example), the conditional expectation $\mathbb{E}(D|W = 0)$ (i.e. the "drift" at zero) does not enter in the first term of the bound.

(3) In our main example (see Section 4), $\mathbb{E}(D^2|W) = 2\lambda$ and so the second error term also vanishes. The third error term $\mathbb{E}|D|^3$ can be bounded using the Cauchy-Schwarz inequality $\mathbb{E}|D|^3 \leq \sqrt{\mathbb{E}|D|^2\mathbb{E}|D|^4}$. The error term which is difficult to bound in practice is the fourth error term, and later in this section we develop suitable tools (see Theorems 2.2 and 2.3).

Before embarking on the proof of Theorem 1.1, we recall the main idea of Stein's method in our context. As observed by Stein [St2], a random variable Z on $[0, \infty)$ is Exp(1) if and only if $\mathbb{E}[f'(Z) - f(Z)] = -f(0^+)$ for all functions f in a large class of functions (whose precise definition we do not need). Here $f(0^+)$ is the limiting value of $f(a)$ as a approaches 0 from the right. Stein's characterization of the exponential distribution motivates the study of the function $f(x)$ solving the equation

$$f'(x) - f(x) = I[x \le t] - (1 - e^{-t}), \qquad x \ge 0.$$

Indeed, for this f one has that

$$\mathbb{P}(W \le t) - \mathbb{P}(Z \le t) = \mathbb{E}[f'(W) - f(W)],$$

and the problem becomes that of bounding $\mathbb{E}[f'(W) - f(W)]$.

We begin with the following lemma.

Lemma 2.1. *For every $t > 0$, the function*

$$f(x) := e^{-(t-x)^+} - e^{-t}, \qquad x \ge 0, \tag{1}$$

(where in (1) *we define the derivative $f'(t) := f'(t^-)$), satisfies the differential equation*

$$f'(x) - f(x) = \mathbb{I}[x \le t] - (1 - e^{-t}), \qquad x \ge 0, \tag{2}$$

and the bounds

$$\|f\|_\infty \le 1, \quad \left\|f'\right\|_\infty \le 1, \quad \sup_{x,y \ge 0} \left|f'(x) - f'(y)\right| \le 1. \tag{3}$$

The second derivative f'', defined for every $x \ne t$, satisfies

$$\sup_{x \ne t} \left|f''(x)\right| \le 1. \tag{4}$$

Proof. Write

$$f(x) = \begin{cases} e^{-t+x} - e^{-t} & \text{if } x \le t, \\ 1 - e^{-t} & \text{if } x > t. \end{cases}$$

Together with the definition of $f'(t)$ this yields

$$f'(x) = \begin{cases} e^{-t+x} & \text{if } x \le t, \\ 0 & \text{if } x > t. \end{cases}$$

Thus, on $x \le t$,

$$f'(x) - f(x) = 1 - (1 - e^{-t})$$

which is (2), and on $x > t$

$$f'(x) - f(x) = 0 - (1 - e^{-t})$$

which again is (2). The bounds (3) and (4) are straightforward; to obtain the last bound in (3) note that f' is non-negative. $\square$

Now we give a proof of Theorem 1.1.

Proof of Theorem 1.1 Using (2) it is clear that we only need to bound $\mathbb{E}(f'(W) - f(W))$. Fix $t > 0$ and let $F(x) := \int_0^x f(y)dy$. By Taylor expansion,

$$\begin{aligned} 0 &= \mathbb{E}\left(F(W') - F(W)\right) \\ &= \mathbb{E}\left(Df(W)\right) + \mathbb{E}\left(D^2 \int_0^1 (1-s) f'(W + sD) ds\right) \\ &= \mathbb{E}\left(Df(W) + \tfrac{1}{2} D^2 f'(W)\right) + \mathbb{E}\left(D^2 J\right) \end{aligned} \tag{5}$$

where

$$J := \int_0^1 (1-s)(f'(W+sD) - f'(W))ds.$$

Let A be the event that $W \wedge W' \leq t \leq W \vee W'$. On A^c we thus have for every $0 \leq s \leq 1$

$$\left|f'(W+sD) - f'(W)\right| \leq s\,|D| \tag{6}$$

by (4), whereas on A we have

$$\left|f'(W+sD) - f'(W)\right| \leq 1 \tag{7}$$

by (3). Dividing (5) by λ and noting that $f(0) = 0$, and thus $f(W) = \mathbb{I}[W > 0]f(W)$, we can use this to obtain that

$$\begin{aligned}
\mathbb{E}(f'(W) - f(W)) &= \mathbb{E}(f'(W) - f(W)) - \frac{1}{\lambda}\left(\mathbb{E}\left(F(W') - F(W)\right)\right) \\
&= -\mathbb{E}\left((\tfrac{1}{\lambda}\mathbb{E}(D|W) + 1)f(W)\mathbb{I}[W > 0]\right) \\
&+ \mathbb{E}\left((1 - \tfrac{1}{2\lambda}\mathbb{E}(D^2|W))f'(W)\right) \\
&- \tfrac{1}{\lambda}\mathbb{E}\left(\mathbb{I}[A^c]D^2J\right) - \tfrac{1}{\lambda}\mathbb{E}\left(\mathbb{I}[A]D^2J\right).
\end{aligned} \tag{8}$$

Invoking the bounds (6) and (7), we have

$$\mathbb{I}[A^c]D^2\,|J| \leq \tfrac{1}{6}|D|^3, \mathbb{I}[A]D^2\,|J| \leq \tfrac{1}{2}D^2\mathbb{I}[|W-t| \leq |D|],$$

where the second inequality uses the fact that A implies $|W - t| \leq |D|$. Combining these bounds with (8) and the bounds $||f||_\infty, ||f'||_\infty \leq 1$ from Lemma 2.1 completes the proof. $\square$

The quantity that is difficult to bound in practice when applying Theorem 1.1 is

$$\mathbb{E}\left(D^2\mathbb{I}[|W-t| \leq |D|]\right).$$

One tool which is useful for bounding this quantity is the following theorem.

Theorem 2.2. *Assume that W and W' are real valued random variables on the same probability space such that $\mathcal{L}(W') = \mathcal{L}(W)$. Let $D = W' - W$. Then for any $t \in \mathbb{R}$ and $c > 0$,*

$$\mathbb{E}\big(D^2\mathbb{I}\{|W-t| \leq |D|\}\big) \leq 4c\mathbb{E}\big|\mathbb{E}(D|W)\big| + \mathbb{E}\big(D^2\mathbb{I}\{|D| > c\}\big).$$

However Theorem 2.2 does not always give good bounds. The next result, though more demanding, can lead to sharper bounds.

Theorem 2.3. *Assume that W and W' are non-negative random variables on the same probability space such that $\mathcal{L}(W') = \mathcal{L}(W)$; let $D = W' - W$. Then for any positive constants t, k_1, k_2, K_1, K_2 and K_3 (where $k_2 < k_1$ and $K_2 < K_3$) we have*

$$\mathbb{E}\left\{D^2\mathbb{I}[|W-t| \leq |D|]\right\} \leq k_2 + k_1e_2 + e_1 + \frac{\mathbb{E}\,|\mathbb{E}(D|W)|}{K_3 - K_2}\times$$
$$\left(k_2 \cdot \ln(k_1/k_2) + \sqrt{32tk_2k_1} + 2K_1t^{1/2}k_1 + 4K_1\sqrt{k_1k_2} + 4K_1(tk_2k_1^3)^{1/4}\right)$$

where

$$\begin{aligned}
e_1 &:= \mathbb{E}\left\{\mathbb{E}(D^2|W)\cdot\mathbb{I}\left[\mathbb{E}(D^2|W) > k_1 \text{ or } \mathbb{E}(D^4|W) > k_2(W+t)\right]\right\} \\
e_2 &:= \mathbb{P}\left[\mathbb{E}(D^2|W) < K_3 \text{ or } \mathbb{E}(D^4|W) > K_1^2K_2W\right]
\end{aligned}$$

The following lemma will be used in the proofs of both Theorems 2.2 and 2.3.

Lemma 2.4. *Suppose that W and W' are random variables on the same probability space such that $\mathcal{L}(W') = \mathcal{L}(W)$; set $D = W' - W$. Then, for any $a \leq b \in \mathbb{R}$ and $K > 0$,*

$$\mathbb{E}\left(D^2 \mathbb{I}[a \leq W \leq b, |D| \leq K]\right) \leq (b - a + 2K)\mathbb{E}\left|\mathbb{E}(D|W)\right|$$

Proof. Define

$$h(x) = \begin{cases} -\frac{1}{2}(b-a) - K & \text{if } x < a - K, \\ x - \frac{1}{2}(a+b) & \text{if } a - K \leq x \leq b + K, \\ \frac{1}{2}(b-a) + K & \text{if } x > b + K. \end{cases}$$

and $H(x) := \int_0^x h(t)dt$. Observe that for any $0 \leq s \leq 1$,

$$\begin{aligned} \mathbb{I}[a \leq W \leq b, |D| \leq K] &\leq \mathbb{I}[a - K \leq W + sD \leq b + K] \\ &= h'(W + sD), \end{aligned}$$

and that

$$\|h\|_\infty = \tfrac{1}{2}(b-a) + K. \tag{9}$$

Using Taylor expansion we have

$$\begin{aligned} 0 &= \mathbb{E}H(W') - \mathbb{E}H(W) \\ &= \mathbb{E}(Dh(W)) + \mathbb{E}\left(D^2 \int_0^1 (1-s)h'(W+sD)ds\right), \end{aligned} \tag{10}$$

and thus

$$\begin{aligned} &\mathbb{E}\left(D^2 \mathbb{I}[a \leq W \leq b, |D| \leq K]\right) \\ &\quad = 2\mathbb{E}\left(D^2 \int_0^1 (1-s)\mathbb{I}[a \leq W \leq b, |D| \leq K]ds\right) \\ &\quad \leq 2\mathbb{E}\left(D^2 \int_0^1 (1-s)h'(W+sD)ds\right) \\ &\quad = -2\mathbb{E}(Dh(W)) \qquad \text{[by (10)]} \\ &\quad \leq 2\left|\mathbb{E}(\mathbb{E}(D|W)h(W))\right| \\ &\quad \leq 2\|h\|_\infty \mathbb{E}\left|\mathbb{E}(D|W)\right| \end{aligned}$$

which together with (9) proves the claim. □

As the following argument shows, Theorem 2.2 is a straightforward consequence of Lemma 2.4.
Proof of Theorem 2.2 Clearly

$$\begin{aligned} \mathbb{E}\big(D^2\mathbb{I}\{|W-t| \leq |D|\}\big) &= \mathbb{E}\big(D^2\mathbb{I}\{|W-t| \leq |D|, |D| > c\}\big) \\ &\quad + \mathbb{E}\big(D^2\mathbb{I}\{|W-t| \leq |D|, |D| \leq c\}\big). \end{aligned}$$

The first term is at most $\mathbb{E}\big(D^2\mathbb{I}\{|D| > c\}\big)$. To upper bound the second term, note that if $|W-t| \leq |D| \leq c$, then $a \leq W \leq b$ where $a = t - c$ and $b = t + c$. Hence Lemma 2.4 gives that

$$\mathbb{E}\big(D^2\mathbb{I}\{|W-t| \leq |D|, |D| \leq c\}\big) \leq 4c\mathbb{E}\big|\mathbb{E}(D|W)\big|.$$

□

We close this section by proving Theorem 2.3.

PROOF OF THEOREM 2.3 Define

$$B(W) := \mathbb{I}\left[\mathbb{E}[(D^2\mathbb{I}[|D| > K_1W^{1/2}])|W] \le K_2, \mathbb{E}(D^2|W) \ge K_3\right].$$

Now note that

$$\mathbb{E}[(D^2\mathbb{I}[D^2 > K_1^2W])|W] \le \frac{\mathbb{E}(D^4|W)}{K_1^2W}.$$

From this it is easy to see that

$$\mathbb{E}(1 - B(W)) \le \mathbb{P}\left[\mathbb{E}(D^2|W) < K_3 \text{ or } \mathbb{E}(D^4|W) > K_1^2K_2W\right] = e_2$$

Note that if $B(W) = 1$ then

$$\begin{aligned}\mathbb{E}[(D^2\mathbb{I}[|D| \le K_1W^{1/2}])|W] &= \mathbb{E}(D^2|W) - \mathbb{E}[(D^2\mathbb{I}[|D| > K_1W^{1/2}])|W]\\ &\ge K_3 - K_2.\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{P}[a \le W \le b] &\le \mathbb{E}\left\{\mathbb{I}[a \le W \le b]B(W)\right\} + e_2\\ &= \mathbb{E}\left\{\frac{K_3 - K_2}{K_3 - K_2}\mathbb{I}[a \le W \le b]B(W)\right\} + e_2\\ &\le \frac{1}{K_3 - K_2}\mathbb{E}\left\{D^2\mathbb{I}[a \le W \le b, |D| \le K_1W^{1/2}]\right\} + e_2\\ &\le \frac{1}{K_3 - K_2}\mathbb{E}\left\{D^2\mathbb{I}[a \le W \le b, |D| \le K_1b^{1/2}]\right\} + e_2\\ &\le \frac{1}{K_3 - K_2}(b - a + 2K_1b^{1/2})\mathbb{E}\left|\mathbb{E}(D|W)\right| + e_2\end{aligned} \tag{11}$$

where the last inequality is due to Lemma 2.4.

Now, define

$$A(W) := \mathbb{I}\left[\mathbb{E}(D^2|W) \le k_1, \mathbb{E}(D^4|W) \le k_2(W + t)\right]$$

Then,

$$\begin{aligned}&\mathbb{E}\{D^2(1 - A(W))\}\\ &\quad = \mathbb{E}\left\{\mathbb{E}(D^2|W) \cdot \mathbb{I}\left[\mathbb{E}(D^2|W) > k_1 \text{ or } \mathbb{E}(D^4|W) > k_2(W + t)\right]\right\}\\ &\quad = e_1.\end{aligned}$$

It follows that

$$\begin{aligned}&\mathbb{E}\left\{D^2\mathbb{I}[|W - t| \le |D|]\right\}\\ &\quad \le \mathbb{E}\left\{D^2\mathbb{I}[(W - t)^2 \le D^2]A(W)\right\} + e_1\\ &\quad \le \mathbb{E}\left\{\min\{D^2, D^4(W - t)^{-2}\}A(W)\right\} + e_1\\ &\quad \le \mathbb{E}\left\{\min\{\mathbb{E}(D^2|W), \mathbb{E}(D^4|W)(W - t)^{-2}\}A(W)\right\} + e_1\\ &\quad \le \mathbb{E}\left\{\min\{k_1, k_2(W + t)(W - t)^{-2}\}\right\} + e_1\end{aligned}$$

Now,

$$\begin{aligned}
&\mathbb{E}\left\{\min\{k_1, k_2(W+t)(W-t)^{-2}\}\right\} \\
&\quad = \int_0^\infty \mathbb{P}[k_1 \geq x, k_2(W+t)(W-t)^{-2} \geq x]dx \\
&\quad = \int_0^{k_1} \mathbb{P}[k_2(W+t)(W-t)^{-2} \geq x]dx \\
&\quad \leq k_2 + \int_{k_2}^{k_1} \mathbb{P}[k_2(W+t)(W-t)^{-2} \geq x]dx
\end{aligned}$$

Suppose that $k_2(W+t)(W-t)^{-2} \geq x$. Then, solving the equation $k_2(w+t)(w-t)^{-2} = x$, one has that

$$\begin{aligned}
W &\in \left[t + \frac{k_2}{2x} - \sqrt{\frac{2tk_2}{x} + \frac{k_2^2}{4x^2}}, t + \frac{k_2}{2x} + \sqrt{\frac{2tk_2}{x} + \frac{k_2^2}{4x^2}}\right] \\
&\subset \left[t - \sqrt{\frac{2tk_2}{x}}, t + \frac{k_2}{x} + \sqrt{\frac{2tk_2}{x}}\right].
\end{aligned}$$

Thus, combining this with the concentration inequality (11),

$$\begin{aligned}
&\mathbb{E}\left\{D^2\mathbb{I}[|W-t| \leq |D|]\right\} \\
&\leq k_2 + e_1 + \int_{k_2}^{k_1} \mathbb{P}[k_2(W+t)(W-t)^{-2} \geq x]dx \\
&\leq k_2 + e_1 + \int_{k_2}^{k_1} \mathbb{P}\left[t - \sqrt{\frac{2tk_2}{x}} \leq W \leq t + \frac{k_2}{x} + \sqrt{\frac{2tk_2}{x}}\right] dx \\
&\leq k_2 + k_1e_2 + e_1 + \frac{\mathbb{E}\left|\mathbb{E}(D|W)\right|}{K_3 - K_2} \int_{k_2}^{k_1} \left(\frac{k_2}{x} + \sqrt{\frac{8tk_2}{x}} + 2K_1\left(t + \frac{k_2}{x} + \sqrt{\frac{2tk_2}{x}}\right)^{1/2}\right) dx \\
&\leq k_2 + k_1e_2 + e_1 + \frac{\mathbb{E}\left|\mathbb{E}(D|W)\right|}{K_3 - K_2} \int_{k_2}^{k_1} \left(\frac{k_2}{x} + \sqrt{\frac{8tk_2}{x}} + 2K_1t^{1/2} + 2K_1\sqrt{\frac{k_2}{x}} + 2K_1\left(\frac{2tk_2}{x}\right)^{1/4}\right) dx \\
&\leq k_2 + k_1e_2 + e_1 + \frac{\mathbb{E}\left|\mathbb{E}(D|W)\right|}{K_3 - K_2} \times \\
&\left(k_2(\ln k_1 - \ln k_2) + \sqrt{32tk_2k_1} + 2K_1t^{1/2}k_1 + 4K_1\sqrt{k_1k_2} + \tfrac{8}{3}K_1(2tk_2)^{1/4}k_1^{3/4}\right).
\end{aligned}$$

This proves the claim. $\square$

3. Berry-Esséen Bound for the Exponential Law: Version 2

A main goal of this section is to prove Theorem 1.2 from the introduction, and to develop tools for analyzing the error terms which appear in it. In particular, the third term can be hard to bound. One way to bound it is to apply Theorem 2.2 from Section 2. Another way it to use the following more demanding result, which is analogous to Theorem 2.3 from Section 2.

Theorem 3.1. *Let W and W' be non-negative random variables on the same probability space such that $\mathcal{L}(W') = \mathcal{L}(W)$. Suppose that $\mathbb{E}(D|W) = -\lambda(W-1)$, where $D = W' - W$ and $\lambda > 0$ is a fixed constant. Then for any $t > 0$ and $\kappa > 0$*

$$\mathbb{E}\big(D^2\mathbb{I}\{|W-t| \le |D|\}\big) \le 16\lambda^2\kappa^2 + 1040\lambda^{3/2}\mathbb{E}|W-1|\kappa\sqrt{t} \\ + 8\lambda\epsilon_2(\tfrac{1}{3}t)t + \epsilon_1(t)$$

where ϵ_1 and ϵ_2 are functions defined on $(0,\infty)$ as

$$\epsilon_1(t) := \mathbb{E}\big[\mathbb{E}(D^2|W)\mathbb{I}\{\mathbb{E}(D^2|W) > 2\lambda(W+t) \text{ or } \mathbb{E}(D^4|W) > 4\lambda^2(\kappa^2W^2 + \kappa^2t^2)\}\big]$$

and

$$\epsilon_2(t) := \mathbb{P}\{\mathbb{E}(D^2|W) < 2\lambda(W - \tfrac{1}{4}t) \text{ or } \mathbb{E}(D^4|W) > 4\lambda^2(\kappa^2W^2 + \kappa^2t^2)\}.$$

Moreover, the above bound holds if the assumption of positivity of W is replaced by the assumption that W is non-negative and assumes only finitely many values.

Remarks:

(1) The idea behind the formulation of Theorem 3.1 is the following: in many problems, we have $\mathbb{E}(D^4|W) \le 4\lambda^2(\kappa^2W^2 + \eta)$ where κ is some constant and η is a negligible term (possibly random).
(2) The random variable W in the example of this paper can assume the value 0 with positive probability.

It is easy to check by integration by parts that if a random variable Z on $[0,\infty)$ is Exp(1), then $\mathbb{E}[Zf'(Z) - (Z-1)f(Z)] = 0$ for well behaved functions f. This motivates the study of the solution $f(x)$ to the equation

$$xf'(x) - (x-1)f(x) = \mathbb{I}\{x \le t\} - (1 - e^{-t}), \quad x \ge 0.$$

Indeed, for such f one has that

$$\mathbb{P}(W \le t) - \mathbb{P}(Z \le t) = \mathbb{E}[Wf'(W) - (W-1)f(W)],$$

and the problem becomes that of bounding

$$\mathbb{E}[Wf'(W) - (W-1)f(W)].$$

Remark: Earlier authors (Mann [Mn], Luk [Lu], Pickett and Reinert [Re]) studied solutions of the equation

$$xf''(x) - (x-1)f'(x) = h(x) - \int_0^\infty e^{-x}h(x),$$

for functions h whose first k derivatives are bounded. This is complementary to our work, since our primary interest is in the function $h(x) = \mathbb{I}\{x \le t\}$, which is not smooth.

Lemma 3.2. *For every $t \in \mathbb{R}$, the function*

$$f(x) := \frac{e^{-(t-x)^+} - e^{-t}}{x}, \quad x > 0$$

satisfies the equation

$$xf'(x) - (x-1)f(x) = \mathbb{I}\{x \le t\} - (1 - e^{-t}), \quad x \in \mathbb{R}^+, \tag{12}$$

where f' denotes the left-hand derivative of f. Moreover, one has the bounds

$$||f'||_\infty \le t^{-1}, \quad ||f''||_\infty \le max\{t^{-1}, 2t^{-2}\}.$$

Proof. Clearly, f is infinitely differentiable on $\mathbb{R}^+\backslash\{t\}$. The left-hand and right-hand derivatives at t exist and are unequal, which is why we let f' denote the left-hand derivative of f. Then for $0 < x \le t$,

$$f'(x) = \frac{d}{dx}\left(\frac{e^{-(t-x)} - e^{-t}}{x}\right) = \frac{xe^x - e^x + 1}{x^2}e^{-t} \tag{13}$$

which gives

$$xf'(x) - (x-1)f(x) = 1 - (1 - e^{-t}).$$

Similarly, for $x > t$,

$$f'(x) = \frac{d}{dx}\left(\frac{1 - e^{-t}}{x}\right) = \frac{e^{-t} - 1}{x^2} \tag{14}$$

which gives

$$xf'(x) - (x-1)f(x) = -(1 - e^{-t}).$$

Thus, the function f is a solution to (12).

The easiest way to get a uniform bound on f' is perhaps by directly expanding in power series. When $0 < x \le t$, we recall (13) to get

$$f'(x) = \frac{xe^x - e^x + 1}{x^2}e^{-t} = e^{-t}\sum_{k=0}^{\infty}\frac{x^k}{k!(k+2)}. \tag{15}$$

This shows that for $x \in (0, t]$,

$$0 \le f'(x) \le e^{-t}\sum_{k=0}^{\infty}\frac{t^k}{(k+1)!} = \frac{1 - e^{-t}}{t} \le \min\{1, t^{-1}\}.$$

Again, for $x > t$, we directly see from (14) that $f'(x) \le 0$ and

$$|f'(x)| \le \frac{1 - e^{-t}}{t^2}$$

Combining, we get

$$\|f'\|_\infty \le t^{-1}.$$

Now, f' is positive in $(0, t]$ and negative in (t, ∞). Therefore f attains its maximum at t. It is now easy to see that for all $x > 0$,

$$0 \le f(x) \le \frac{1 - e^{-t}}{t} \le 1.$$

Using (15) we see that for $0 < x \le t$

$$0 \le f''(x) = e^{-t}\sum_{k=1}^{\infty}\frac{kx^{k-1}}{k!(k+2)} = e^{-t}\sum_{k=0}^{\infty}\frac{x^k}{k!(k+3)} \le \frac{1 - e^{-t}}{t}$$

and for $x > t$,

$$0 \le f''(x) = \frac{2(1 - e^{-t})}{x^3} \le \frac{2(1 - e^{-t})}{t^3}.$$

Combining, we get, for all $x > 0$,

$$0 \le f''(x) \le \max\{t^{-1}, 2t^{-2}\}.$$

This completes the proof. $\square$

Now the main results of this section will be proved.

PROOF OF THEOREM 1.2 Fix $t > 0$ and consider the Stein equation

$$xf'(x) - (x-1)f(x) = \mathbb{I}[x \le t] - \mathbb{P}[Z \le t] \tag{16}$$

for $x > 0$, where $Z \sim \mathrm{Exp}(1)$. From Lemma 3.2, its solution f satisfies the non-uniform bounds

$$\|f'\|_\infty \le t^{-1},\ \|f''\|_\infty \le \max\{t^{-1}, 2t^{-2}\}, \tag{17}$$

where f'' denotes the left derivative of f', as f' has a discontinuity in t.

Assume first that W is positive. Defining $G(w) = \int_0^w f(x)dx$, Taylor's expansion gives that

$$G(W') = G(W) + Df(W) + D^2 \int_0^1 (1-s) f'(W+sD)\, ds.$$

The hypothesis $\mathbb{E}(D|W) = -\lambda(W-1)$ gives that

$$\begin{aligned} 0 &= \mathbb{E}\left\{G(W') - G(W)\right\} \\ &= \mathbb{E}\left\{-\lambda(W-1)f(W)\right\} + \mathbb{E}\left\{D^2 \int_0^1 (1-s)f'(W+sD)\, ds\right\}, \end{aligned}$$

and hence

$$\mathbb{E}\left\{(W-1)f(W)\right\} = \mathbb{E}\left\{\frac{1}{\lambda} D^2 \int_0^1 (1-s)f'(W+sD)\, ds\right\}.$$

Taking expectation on (16) with respect to W, we thus have

$$\begin{aligned} &\mathbb{P}[W \le t] - \mathbb{P}[Z \le t] \\ &\quad = \mathbb{E}\left\{Wf'(W) - (W-1)f(W)\right\} \\ &\quad = \mathbb{E}\left\{Wf'(W) - \frac{1}{\lambda} D^2 \int_0^1 (1-s)f'(W+sD)\, ds\right\} \\ &\quad = \mathbb{E}\left\{\left(W - D^2/(2\lambda)\right) f'(W)\right\} \\ &\qquad\quad + \mathbb{E}\left\{\frac{D^2}{\lambda}\left[\frac{1}{2} f'(W) - \int_0^1 (1-s) f'(W+sD)\, ds\right]\right\} \\ &\quad = \mathbb{E}\left\{\left(W - D^2/(2\lambda)\right) f'(W)\right\} \\ &\qquad\quad + \mathbb{E}\left\{\frac{D^2}{\lambda} \int_0^1 (1-s)(f'(W) - f'(W+sD))ds\right\}. \end{aligned} \tag{18}$$

Note now that for any $x, y > 0$,

$$\left|f'(x) - f'(y)\right| \le \begin{cases} \|f''\|_\infty |x-y| & \text{if } x \text{ and } y \text{ lie on the same side of } t, \\ 2\|f'\|_\infty & \text{otherwise.} \end{cases}$$

Also, if x and y lie on different sides of t, then $|x-t| \le |x-y|$. Thus

$$\begin{aligned}
&\int_0^1 (1-s)\left|f'(W) - f'(W+sD)\right| ds \\
&\le \left\|f''\right\|_\infty \int_0^1 (1-s)s|D|\,ds + 2\left\|f'\right\|_\infty \int_0^1 (1-s)\mathbb{I}\left[|W-t| \le |sD|\right] ds \\
&\le \frac{1}{6}|D|\left\|f''\right\|_\infty + \left\|f'\right\|_\infty \mathbb{I}\left[|W-t| \le |D|\right].
\end{aligned}$$

Putting the steps together we obtain from (18)

$$\begin{aligned}
|\mathbb{P}[W \le t] - \mathbb{P}[Z \le t]| \;\le\;& \left\|f'\right\|_\infty \mathbb{E}\left|W - \frac{\mathbb{E}(D^2|W)}{2\lambda}\right| \\
&+\frac{1}{6\lambda}\left\|f''\right\|_\infty \mathbb{E}|D|^3 + \frac{1}{\lambda}\left\|f'\right\|_\infty \mathbb{E}\left\{D^2\mathbb{I}[|W-t| \le |D|]\right\},
\end{aligned}$$

and with the bounds (17) the claim follows for positive W.

To treat the case where W can also equal 0, choose $0 < \delta < 1$ and define $W_\delta := (1-\delta)W + \delta$, $W'_\delta := (1-\delta)W' + \delta$ and $t_\delta := (1-\delta)t + \delta$. One sees that W_δ is a positive random variable, and that $\mathbb{E}(D_\delta|W_\delta) = -\lambda(W_\delta - 1)$ where λ is the same as for the pair (W, W'). Moreover $\mathbb{P}\{W \le t\} = \mathbb{P}\{W_\delta \le t_\delta\}$, so it follows that

$$\begin{aligned}
|\mathbb{P}\{W \le t\} - \mathbb{P}\{Z \le t\}| \le& \frac{1}{2\lambda t_\delta}\mathbb{E}\big|2\lambda W_\delta - \mathbb{E}(D_\delta^2|W)\big| + \frac{\max\{t_\delta^{-1}, 2t_\delta^{-2}\}}{4\lambda}\mathbb{E}|D_\delta|^3 \\
&+ \frac{1}{\lambda t_\delta}\mathbb{E}\big(D_\delta^2\mathbb{I}\{|W_\delta - t_\delta| \le |D_\delta|\}\big).
\end{aligned}$$

Since $D_\delta = (1-\delta)D$, the first two error terms are continuous in δ and converge to the corresponding error terms for W when $\delta \to 0$. The same is true for the third error term, as can be seen from the fact that $|W_\delta - t_\delta| \le |D_\delta|$ if and only if $|W-t| \le |D|$. This completes the proof. □

Next, we prove Theorem 3.1.

Proof of Theorem 3.1 First we treat the case that W is always positive. Throughout we shall be using $V := (2\lambda)^{-1/2}(W' - W)$ instead of $D(= W' - W)$, simply because D occurs with a factor of $(2\lambda)^{-1/2}$ attached with it on most occasions.

Suppose for each $0 < s \le t$, we have numbers $u(s,t)$ and $v(s,t)$ such that whenever $s \le a \le b \le t$, we have

$$\mathbb{P}\{a \le W \le b\} \le u(s,t)(b-a) + v(s,t).$$

Fix $t \in \mathbb{R}$. Let $A(W) = \mathbb{I}\{\mathbb{E}(V^2|W) \le W + t,\ \mathbb{E}(V^4|W) \le \kappa^2(W+t)^2\}$. Then

$$\begin{aligned}
&\mathbb{E}\big(V^2(1 - A(W))\big) \\
&\le \mathbb{E}\big(\mathbb{E}(V^2|W)\mathbb{I}\{\mathbb{E}(V^2|W) > W + t \text{ or } \mathbb{E}(V^4|W) > \kappa^2W^2 + \kappa^2t^2\}\big) \\
&=: e_1(t).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbb{E}(V^2\mathbb{I}\{|W-t| \le |W'-W|\}) \\
&\le \mathbb{E}(V^2\mathbb{I}\{(W'-W)^2 \ge (W-t)^2\}A(W)) + e_1(t) \\
&\le \mathbb{E}(\min\{2\lambda(W-t)^{-2}V^4, V^2\}A(W)) + e_1(t) \\
&\le \mathbb{E}(\min\{2\lambda(W-t)^{-2}\mathbb{E}(V^4|W), \mathbb{E}(V^2|W)\}A(W)) + e_1(t) \\
&\le \mathbb{E}(\min\{2\lambda\kappa^2(W-t)^{-2}(W+t)^2, W+t\}) + e_1(t).
\end{aligned}$$

Now

$$\begin{aligned}
&\mathbb{E}(\min\{2\lambda\kappa^2(W-t)^{-2}(W+t)^2, W+t\}) \\
&= \int_0^\infty \mathbb{P}\{2\lambda\kappa^2(W-t)^{-2}(W+t)^2 \ge x, W+t \ge x\}\, dx.
\end{aligned} \tag{19}$$

Now take any $x \ge 8\lambda\kappa^2$. Let $c(x) = \sqrt{\frac{2\lambda\kappa^2}{x}}$. Then the following are easily seen to be equivalent:

$$\begin{aligned}
2\lambda\kappa^2(W-t)^{-2}(W+t)^2 \ge x &\iff |W-t| \le c(x)(W+t) \\
&\iff \frac{1-c(x)}{1+c(x)}t \le W \le \frac{1+c(x)}{1-c(x)}t.
\end{aligned}$$

Let $a(x) = (1-c(x))/(1+c(x))$ and $b(x) = (1+c(x))/(1-c(x))$. Note that since $x \ge 8\lambda\kappa^2$, therefore $c(x) \le 1/2$ and so $a(x) \ge 1/3$, $b(x) \le 3$, and

$$b(x) - a(x) = \frac{4c(x)}{1-c(x)^2} \le \tfrac{16}{3}c(x).$$

Now if $W \le 3t$ then $W+t \le 4t$. Thus, the integrand in (19) is zero for $x > 4t$. Combining, we see that

$$\begin{aligned}
&\mathbb{E}(\min\{2\lambda(W-t)^{-2}(W+t)^2, W+t\}) \\
&\le 8\lambda\kappa^2 + \int_{8\lambda\kappa^2}^{4t} \mathbb{P}\{a(x)t \le W \le b(x)t\}dx \\
&\le 8\lambda\kappa^2 + \int_{8\lambda\kappa^2}^{4t} \big(u(\tfrac{1}{3}t, 3t)\tfrac{16}{3}c(x)t + v(\tfrac{1}{3}t, 3t)\big)dx \\
&\le 8\lambda\kappa^2 + 22u(\tfrac{1}{3}t, 3t)t\kappa\sqrt{2\lambda t} + 4tv(\tfrac{1}{3}t, 3t).
\end{aligned} \tag{20}$$

Next, we proceed to find suitable values of $u(s,t)$ and $v(s,t)$. Fix $0 < s \le a \le b \le t$. Let

$$B(W) = \mathbb{I}\{\mathbb{E}(V^2\mathbb{I}\{|V| > 2\kappa\sqrt{W+a}\}|W) \le \tfrac{1}{4}(W+a),\ \mathbb{E}(V^2|W) \ge W - \tfrac{1}{4}a\}.$$

Now note that

$$\mathbb{E}(V^2\mathbb{I}\{|V| > 2\kappa\sqrt{W+a}\}|W) \le \frac{\mathbb{E}(V^4|W)}{4\kappa^2(W+a)}.$$

From this it is easy to see that

$$\begin{aligned}
\mathbb{E}(1-B(W)) &\le \mathbb{P}\{\mathbb{E}(V^2|W) < W - \tfrac{1}{4}a \text{ or } \mathbb{E}(V^4|W) > \kappa^2W^2 + \kappa^2a^2\} \\
&=: e_2(a)
\end{aligned}$$

Note that if $B(W)=1$ then $\mathbb{E}(V^2\mathbb{I}\{|V|\le 2\kappa\sqrt{W+a}\}|W)\ge \frac{3}{4}W-\frac{1}{2}a$. So, if $W\ge a$ and $B(W)=1$, $\mathbb{E}(V^2\mathbb{I}\{|V|\le 2\kappa\sqrt{W+a}\}|W)\ge \frac{1}{4}a$. Thus,

$$\begin{aligned} a\mathbb{P}\{a\le W\le b\} &\le 4\mathbb{E}\big(V^2\mathbb{I}\{a\le W\le b,\ |V|\le 2\kappa\sqrt{W+a}\}\big)+ae_2(a)\\ &\le 4\mathbb{E}\big(V^2\mathbb{I}\{a\le W\le b,\ |V|\le 2\kappa\sqrt{b+a}\}\big)+ae_2(a)\\ &= 2\lambda^{-1}\mathbb{E}\big(D^2\mathbb{I}\{a\le W\le b,\ |D|\le 2\kappa\sqrt{2\lambda(b+a)}\}\big)+ae_2(a). \end{aligned}$$

where $D=W'-W$. Using Lemma 2.4, we get

$$a\mathbb{P}\{a\le W\le b\}\le 2\big(b-a+4\kappa\sqrt{2\lambda(b+a)}\big)\mathbb{E}|W-1|+ae_2(a).$$

Finally, note that e_2 is a monotonically decreasing function. Thus, we can take

$$u(s,t)=\frac{2\mathbb{E}|W-1|}{s}$$

and

$$v(s,t)=\frac{16\kappa\sqrt{\lambda t}\mathbb{E}|W-1|}{s}+e_2(s).$$

Using these expressions for u and v in (20), we get

$$\mathbb{E}\big(V^2\mathbb{I}\{|W-t|\le|W'-W|\}\big)\le 8\lambda\kappa^2+520\mathbb{E}|W-1|\kappa\sqrt{\lambda t}+4e_2(\tfrac{1}{3}t)t+e_1(t).$$

Put $\epsilon_1(t)=2\lambda e_1(t)$ and $\epsilon_2(t)=e_2(t)$ to get the final expression in Theorem 3.1.

Finally, suppose that W might take the value 0, but that W assumes only finitely many values. As in the proof of Theorem 1.2, for $0<\delta<1$ define $W_\delta:=(1-\delta)W+\delta$, $W'_\delta:=(1-\delta)W'+\delta$ and $t_\delta:=(1-\delta)t+\delta$. Since $|W_\delta-t_\delta|\le|D_\delta|$ if and only if $|W-t|\le|D|$, it follows that $\mathbb{E}(D^2\mathbb{I}\{|W-t|\le|D|\})$ is the limit as $\delta\to 0$ of $\mathbb{E}(D_\delta^2\mathbb{I}\{|W_\delta-t_\delta|\le|D_\delta|\})$. It is easily checked that $\mathbb{E}(D_\delta^2|W)>2\lambda(W_\delta+t_\delta)$ implies that $\mathbb{E}(D^2|W)>2\lambda(W+t)$ and that $\mathbb{E}(D_\delta^4|W)>4\lambda^2(\kappa^2W_\delta^2+\kappa^2t_\delta^2)$ implies that $\mathbb{E}(D^4|W)>4\lambda^2(\kappa^2W^2+\kappa^2t^2)$. We claim that $\mathbb{E}(D_\delta^2|W)<2\lambda(W_\delta-\frac{1}{4}t_\delta)$ implies that $\mathbb{E}(D^2|W)<2\lambda(W-\frac{1}{4}t)$ provided that δ is sufficiently small. Indeed, since W takes only finitely many values, there is an $m_t>0$ such that $\mathbb{E}(D^2|W)<2\lambda(W-\frac{1}{4}t)$ if and only if $\mathbb{E}(D^2|W)<2\lambda(W-\frac{1}{4}t)+m_t$. The claim now follows since $\mathbb{E}(D_\delta^2|W)<2\lambda(W_\delta-\frac{1}{4}t_\delta)$ implies that $\mathbb{E}(D^2|W)<2\lambda(W-\frac{1}{4}t)+\frac{3\lambda\delta}{2(1-\delta)}+\delta\mathbb{E}(D^2|W)$. Hence the theorem follows by letting $\delta\to 0$. □

4. Example: Spectrum of Bernoulli-Laplace chain

This section proves Theorem 1.4 of the introduction. Throughout we let W denote the random variable defined by

$$W(i):=\frac{(n-2i)(n+2-2i)}{2n},$$

where n is even and $i\in\{0,1,\cdots,\frac{n}{2}\}$ is chosen with probability $\pi(i)$ equal to

$$\frac{\binom{n}{i}-\binom{n}{i-1}}{\binom{n}{n/2}}\quad\text{if }\ 1\le i\le\tfrac{n}{2},\qquad \frac{1}{\binom{n}{n/2}}\quad\text{if } i=0.$$

Letting Z be an Exp(1) random variable and C a universal constant, the upper bound

$$|\mathbb{P}(W\le t)-\mathbb{P}(Z\le t)|\le\frac{C}{\sqrt{n}}$$

will be proved in two steps. Subsection 4.1 uses the machinery of Section 2 to treat the case that $t \leq 1$, and Subsection 4.2 uses the machinery of Section 3 to treat the case that $t \geq 1$. One interesting feature of the proof is that the exchangeable pairs used in these two subsections are different (but closely related). We also show (in Subsection 4.1), that combining the machinery of Section 2 with a concentration inequality, one can obtain, with less effort, a slightly weaker $O(\frac{\log(n)}{\sqrt{n}})$ upper bound.

Finally, Subsection 4.3 shows that the $O(n^{-1/2})$ rate is sharp, by constructing a sequence of n's tending to infinity and corresponding t_n's such that

$$|\mathbb{P}(W_n \leq t_n) - \mathbb{P}(Z \leq t_n)| = \frac{2e^{-2}}{\sqrt{n}} + O(1/n).$$

4.1. **Upper bound for small t.** The purpose of this subsection is to use the machinery of Section 2 to prove Proposition 4.1, which implies the upper bound of Theorem 1.4 of the introduction for $t \leq 1$.

Proposition 4.1.

$$|\mathbb{P}(W \leq t) - \mathbb{P}(Z \leq t)| \leq \frac{C \cdot \max(1, t^{1/2})}{\sqrt{n}}$$

for a universal constant C.

To begin we define an exchangeable pair (W, W') and perform some computations with it. The definition of (W, W') and the fact that the computations work out so neatly may seem unmotivated. There is an algebraic motivation for our choices, and so as not to interrupt our self-contained probabilistic treatment, we explain this in the appendix.

To construct an exchangeable pair (W, W'), we specify a Markov chain K on the set $\{0, 1, \cdots, \frac{n}{2}\}$ which is reversible with respect to π. This means that $\pi(i)K(i,j) = \pi(j)K(j,i)$ for all i, j. Given such a Markov chain K, one obtains the pair (W, W') in the usual way (see for instance [RR]): choose i from π, let $W = W(i)$, and let $W' = W(j)$, where j is obtained from i by taking one step using the Markov chain K.

The Markov chain which turns out to be useful is a birth-death chain on $\{0, 1, \cdots, \frac{n}{2}\}$ where the transition probabilities are

$$K(i, i+1) := \frac{n - i + 1}{n(n-2i)(n-2i+1)}$$
$$K(i, i-1) := \frac{i}{n(n-2i+1)(n-2i+2)}$$
$$K(i,i) := 1 - K(i, i+1) - K(i, i-1),$$

with the exception of $K(i, i+1)$ if $i = n/2$, which we define to be zero.

It is easily checked that K is reversible with respect to π, so the resulting pair (W, W') is exchangeable. (In fact the machinery of Section 2 only uses that W and W' have the same law, which follows from the fact that K has π as a stationary distribution, but the exchangeability is good to record).

Lemma 4.2 performs some moment computations related to the pair (W, W').

Lemma 4.2. *Letting $D := W' - W$, one has that:*

(1) $\mathbb{E}(D|W) = -\frac{2}{n^2}$ *if* $W \neq 0$; $\mathbb{E}(D|W = 0) = \frac{1}{n}$.

(2) $\mathbb{E}(W) = 1$.
(3) $\mathbb{E}(D^2|W) = \frac{4}{n^2}$.
(4) $\mathbb{E}(D^4|W) = \left(\frac{32}{n^3} - \frac{64}{n^4}\right) W + \frac{64}{n^4}$.
(5) $\mathbb{E}(D^4) = \frac{32}{n^3}$.

Proof. Since i is determined by $W(i)$, conditional expectations given W can be computed using conditional expectations given i. Supposing that $i \neq n/2$,

$$\begin{aligned}
\mathbb{E}(D|i) &= K(i,i+1)(W(i+1) - W(i)) + K(i,i-1)(W(i-1) - W(i)) \\
&= \frac{n-i+1}{n(n-2i)(n-2i+1)} \frac{2(2i-n)}{n} + \frac{i}{n(n-2i+1)(n-2i+2)} \frac{2(n-2i+2)}{n} \\
&= -\frac{2}{n^2}.
\end{aligned}$$

If $i = n/2$, then $\mathbb{E}(D|i) = K(i,i-1)(W(i-1) - W(i)) = \frac{1}{n}$, so part 1 is proved.

For part 2, argue as in part 1 (separately treating the cases $i \neq n/2$ and $i = n/2$) to compute that $\mathbb{E}(D^3|W) = -\frac{16}{n^3}(W-1)$. Since W and W' are exchangeable, $\mathbb{E}(D^3) = 0$. Thus

$$\mathbb{E}(W-1) = -\frac{n^3}{16}\mathbb{E}[\mathbb{E}(D^3|W)] = -\frac{n^3}{16}\mathbb{E}(D^3) = 0,$$

so $\mathbb{E}(W) = 1$.

For parts 3 and 4, one argues as in part 1 to compute both sides (separately treating the cases $i \neq n/2$ and $i = n/2$) and checks that they are equal. For part 5, note that

$$\mathbb{E}(D^4) = \mathbb{E}[\mathbb{E}(D^4|W)] = \left(\frac{32}{n^3} - \frac{64}{n^4}\right)\mathbb{E}[W] + \frac{64}{n^4} = \frac{32}{n^3},$$

where the final equality is part 2. $\square$

Using these moment computations, we deduce Proposition 4.1.

Proof of Proposition 4.1 We apply Theorem 1.1 to the pair (W, W') with the value $\lambda = \frac{2}{n^2}$. Then the first two error terms actually vanish. Indeed, part 1 of Lemma 4.2 gives that

$$\mathbb{E}\left|(\lambda^{-1}\mathbb{E}(D|W) + 1)\mathbb{I}[W > 0]\right| = 0,$$

and part 3 of Lemma 4.2 gives that

$$\mathbb{E}\left|\tfrac{1}{2\lambda}\mathbb{E}(D^2|W) - 1\right| = 0.$$

To analyze the third error term, use the Cauchy-Schwarz inequality and parts 3 and 4 of Lemma 4.2 to obtain that

$$\frac{n^2}{12}\mathbb{E}|D^3| \leq \frac{n^2}{12}\sqrt{\mathbb{E}(D^2)\mathbb{E}(D^4)} = \sqrt{\frac{8}{9n}}.$$

To bound the fourth error term, apply Theorem 2.3 with

$$k_1 = \frac{4}{n^2},\ k_2 = \frac{48}{n^3},\ K_1 = \sqrt{\frac{48}{n}},\ K_2 = \frac{1}{n^2},\ K_3 = \frac{4}{n^2}.$$

Note (as required by the theorem), that $K_2 < K_3$ and that for $n > 12$, $k_2 < k_1$. From part 1 of Lemma 4.2 and the fact that $\mathbb{P}(W = 0) = \frac{2}{n+2}$, one computes that $\mathbb{E}|\mathbb{E}(D|W)| = \frac{4}{n(n+2)}$.

It is necessary to upper bound

$$e_1 = \mathbb{E}\left\{\mathbb{E}(D^2|W) \cdot \mathbb{I}\left[\mathbb{E}(D^2|W) > k_1 \text{ or } \mathbb{E}(D^4|W) > k_2(W+t)\right]\right\}.$$

Note from part 3 of Lemma 4.2 that

$$\mathbb{P}\left[\mathbb{E}(D^2|W) > \frac{4}{n^2}\right] = 0$$

and from part 4 of Lemma 4.2 that

$$\begin{aligned} \mathbb{P}\left[\mathbb{E}(D^4|W) > \frac{48}{n^3}(W+t)\right] &\leq \mathbb{P}\left[\mathbb{E}(D^4|W) > \frac{48}{n^3}W\right] \\ &= \mathbb{P}(W < 4/(n+4)) \\ &= \mathbb{P}(W = 0) \\ &= \frac{2}{n+2}. \end{aligned} \tag{21}$$

Thus

$$e_1 \leq \frac{4}{n^2}\mathbb{P}(W=0) = \frac{8}{n^2(n+2)}.$$

It is also necessary to upper bound

$$e_2 = \mathbb{P}\left[\mathbb{E}(D^2|W) < K_3 \text{ or } \mathbb{E}(D^4|W) > K_1^2 K_2 W\right].$$

Note from part 3 of Lemma 4.2 that

$$\mathbb{P}\left[\mathbb{E}(D^2|W) < \frac{4}{n^2}\right] = 0$$

and from (21) that $\mathbb{P}[\mathbb{E}(D^4|W) > \frac{48}{n^3}W] = \frac{2}{n+2}$. Thus $e_2 = \frac{2}{n+2}$. Plugging into Theorem 2.3, one obtains that

$$\frac{n^2}{4}\mathbb{E}\big(D^2\mathbb{I}\{|W-t| \leq |D|\}\big) \leq \frac{C \cdot \max\{1, t^{1/2}\}}{\sqrt{n}},$$

for a universal constant C. This completes the proof. □

To close this subsection, we show how the machinery of Section 2, together with a concentration inequality for $W' - W$, leads to a simpler proof (avoiding the use of Theorem 2.3) that

$$|\mathbb{P}(W \leq t) - \mathbb{P}(Z \leq t)| \leq C\sqrt{\frac{\log(n)}{n}},$$

for a universal constant C. We hope that this approach will be useful in other settings (a concentration inequality for $W' - W$ can be very useful for normal approximation by Stein's method; see the survey [CS]).

The following lemma is helpful for obtaining a concentration result for $W' - W$.

Lemma 4.3. *Let a be an integer such that $0 \leq a \leq \frac{n}{2}$. Then $\binom{n}{\frac{n}{2}-a}/\binom{n}{\frac{n}{2}} \leq e^{-\frac{a(a-1)}{n}}$.*

Proof. The result is visibly true for $a = 0$, so suppose that $a \geq 1$. Observe that

$$\begin{aligned}
\frac{\binom{n}{\frac{n}{2}-a}}{\binom{n}{\frac{n}{2}}} &= \frac{(\frac{n}{2})\cdots(\frac{n}{2}-a+1)}{(\frac{n}{2}+1)\cdots(\frac{n}{2}+a)} \\
&\leq \frac{(\frac{n}{2})\cdots(\frac{n}{2}-a+1)}{(\frac{n}{2})^a} \\
&= \prod_{i=1}^{a-1}(1-\frac{2i}{n}) \\
&= e^{\sum_{i=1}^{a-1}\log(1-\frac{2i}{n})} \\
&\leq e^{-\sum_{i=1}^{a-1}\frac{2i}{n}} \\
&= e^{-\frac{a(a-1)}{n}}.
\end{aligned}$$

$\square$

Proposition 4.4 gives the concentration inequality for $W' - W$. As usual $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

Proposition 4.4. $\mathbb{P}(|W'-W| > c) \leq n^{-5/2}$ *for* $c = \frac{4}{n}\left(\left\lceil\sqrt{\frac{5}{2}n\log(n)}\right\rceil + 1\right)$.

Proof. Since the Markov chain K used to construct (W, W') is a birth death chain, it is easily checked from the definition of W that $|W'(i) - W(i)| \leq \frac{2}{n}(n - 2i + 2)$ for all i. Thus for c as in the proposition,

$$\begin{aligned}
\mathbb{P}(|W'-W| > c) &\leq \mathbb{P}\left(\frac{2}{n}(n-2i+2) > c\right) \\
&= \mathbb{P}\left(i < \frac{n}{2} - \frac{cn}{4} + 1\right) \\
&= \mathbb{P}\left[i < \frac{n}{2} + 1 - \left(\left\lceil\sqrt{\frac{5}{2}n\log(n)}\right\rceil + 1\right)\right].
\end{aligned}$$

From the definition of the probability measure π, it is clear that for integral a, $\mathbb{P}(i < \frac{n}{2} + 1 - a) = \frac{\binom{n}{\frac{n}{2}-a}}{\binom{n}{\frac{n}{2}}}$. Hence the proposition follows from Lemma 4.3. $\square$

This leads to the following proposition.

Proposition 4.5.

$$|\mathbb{P}(W \leq t) - \mathbb{P}(Z \leq t)| \leq C\sqrt{\frac{\log(n)}{n}},$$

for a universal constant C.

Proof. As in the proof of Proposition 4.1, apply Theorem 1.1 to the pair (W, W') with the value $\lambda = \frac{2}{n^2}$. The first three terms are bounded as in the proof of Proposition 4.1. To bound the fourth term, note from Theorem 2.2 that

$$\frac{1}{2\lambda}\mathbb{E}\big(D^2\mathbb{I}\{|W-t| \leq |D|\}\big) \leq n^2 c\mathbb{E}\big|\mathbb{E}(D|W)\big| + \frac{n^2}{4}\mathbb{E}\big(D^2\mathbb{I}\{|D| > c\}\big)$$

for any $c > 0$. From part 1 of Lemma 4.2 one computes that $\mathbb{E}|\mathbb{E}(D|W)| = \frac{4}{n(n+2)}$. One checks from the definitions that $|W' - W| \le 2 + \frac{4}{n}$, so that $(W' - W)^2 \le 16$ since n is even. Choosing $c = \frac{4}{n}\left(\left\lceil \sqrt{\frac{5}{2} n \log(n)} \right\rceil + 1\right)$, it follows from Proposition 4.4 that

$$\mathbb{E}\big(D^2 \mathbb{I}\{|D| > c\}\big) \le 16\mathbb{P}(|D| > c) \le 16 n^{-5/2}.$$

This proves the proposition. □

4.2. **Upper bound for large t.** The purpose of this subsection is to apply the machinery of Section 3 to prove the following Proposition, which gives the upper bound in Theorem 1.4 in the introduction for $t \ge 1$.

Proposition 4.6.

$$|\mathbb{P}(W \le t) - \mathbb{P}(Z \le t)| \le \frac{C \cdot \max(1, t^{-3})}{\sqrt{n}}$$

for a universal constant C.

The pair (W, W') used in this subsection is somewhat different from the pair used in Subsection 4.1; for a discussion of the relationship between the two pairs see the remark below. As with the pair from Subsection 4.1, the definition and the fact that the computations work out so nicely may seem unmotivated. The algebraic motivation for the choices is discussed in the appendix.

To construct an exchangeable pair (W, W'), we specify a Markov chain K on the set $\{0, 1, \cdots, \frac{n}{2}\}$ which is reversible with respect to π (i.e. one has that $\pi(i)K(i,j) = \pi(j)K(j,i)$ for all i, j). Given such a Markov chain K, one obtains the pair (W, W') by choosing i from π, letting $W = W(i)$, and setting $W' = W(j)$, where j is obtained from i by taking one step using the Markov chain K.

The Markov chain which turns out to be useful is a birth-death chain on $\{0, 1, \cdots, \frac{n}{2}\}$ whose only non-zero transition probabilities are

$$K(i, i+1) := \frac{(n-i+1)(n-2i)}{n(n-2i+1)}, \ K(i, i-1) := \frac{i(n-2i+2)}{n(n-2i+1)}.$$

It is easily checked that K is reversible with respect to π so that (W, W') is exchangeable. (In fact the machinery of Section 3 only uses that W and W' have the same law).

Remark: If $K(i,j)$ denotes the transition probabilities of this subsection, and $\tilde{K}(i,j)$ denotes the transition probabilities from Subsection 4.1, one can verify the relation

$$\tilde{K}(i,j) = \frac{4}{n^2} \frac{K(i,j)}{(W(i) - W(j))^2}, \quad \forall i \ne j.$$

Letting $D = W' - W$ for the pair of this subsection and $\tilde{D}, \tilde{W}$ the corresponding quantities for the pair from Subsection 4.1, it follows that

$$\mathbb{E}[\tilde{D}^r | \tilde{W}] = \frac{4}{n^2} \mathbb{E}[D^{r-2} | W]$$

for all r.

Lemma 4.7 performs some moment computations related to the pair (W, W').

Lemma 4.7. *Letting $D := W' - W$, one has that:*

(1) $\mathbb{E}(D|W) = -\frac{4}{n}(W - 1)$.

(2) $\mathbb{E}(W) = 1$.
(3) $\mathbb{E}(D^2|W) = \frac{8}{n}W - \frac{16}{n^2}(W-1)$.
(4) $Var(W) = 1$.
(5) $\mathbb{E}[D^4|W] = \frac{32}{n^2}\left(2W^2 + \frac{12W-8W^2}{n} + \frac{8(1-W)}{n^2}\right)$.
(6) $\mathbb{E}[D^4|W] \leq \frac{256}{n^2}W^2 + \frac{256}{n^4}\mathbb{I}\{W=0\}$.

Proof. For part 1, by the construction of (W, W') one has that

$$\begin{aligned} \mathbb{E}(D|i) &= K(i,i+1)[W(i+1)-W(i)] + K(i,i-1)[W(i-1)-W(i)] \\ &= \frac{(n+1-i)(n-2i)}{n(n+1-2i)}\left(\frac{4i}{n}-2\right) + \frac{i(n-2i+2)}{n(n+1-2i)}\left(2-\frac{4(i-1)}{n}\right). \end{aligned}$$

Elementary simplifications show that this to equal $-\frac{4}{n}(W(i)-1)$.

For part 2, since W and W' have the same law, one has that $\mathbb{E}(D) = 0$. By part 1,

$$\mathbb{E}(D) = \mathbb{E}[\mathbb{E}(D|W)] = -\frac{4}{n}\mathbb{E}(W-1),$$

and the result follows.

For part 3, the construction of (W, W') gives that

$$\begin{aligned} \mathbb{E}[D^2|i] &= K(i,i+1)[W(i+1)-W(i)]^2 + K(i,i-1)[W(i-1)-W(i)]^2 \\ &= \frac{(n+1-i)(n-2i)\left(\frac{4i}{n}-2\right)^2}{n(n+1-2i)} + \frac{i(n-2i+2)\left(2-\frac{4(i-1)}{n}\right)^2}{n(n+1-2i)}. \end{aligned}$$

Part 3 now follows by elementary algebra.

For part 4, observe that

$$\begin{aligned} \mathbb{E}[D^2] &= \mathbb{E}[\mathbb{E}[(W'-W)^2|W]] \\ &= \mathbb{E}[(W')^2] + \mathbb{E}(W^2) - \mathbb{E}[2W\mathbb{E}(W'|W)] \\ &= 2\mathbb{E}(W^2) - \mathbb{E}[2W\mathbb{E}(W'|W)] \\ &= 2\mathbb{E}(W^2) - \mathbb{E}\left[2W\left((1-\frac{4}{n})W + \frac{4}{n}\right)\right] \\ &= \frac{8}{n}\mathbb{E}(W^2) - \frac{8}{n}. \end{aligned}$$

The third equality used that W and W' have the same distribution. The fourth equality used part 1, and the final equality used part 2. Now parts 2 and 3 imply that $\mathbb{E}[D^2] = \frac{8}{n}$. Thus $\mathbb{E}(W^2) = 2$, which together with part 2 implies that $Var(W) = 1$.

For part 5, note by the construction of (W, W') that

$$\begin{aligned} \mathbb{E}(D^4|i) &= K(i,i+1)[W(i+1)-W(i)]^4 + K(i,i-1)[W(i-1)-W(i)]^4 \\ &= \frac{(n+1-i)(n-2i)\left(\frac{4i}{n}-2\right)^4}{n(n+1-2i)} + \frac{i(n-2i+2)\left(2-\frac{4(i-1)}{n}\right)^4}{n(n+1-2i)}. \end{aligned}$$

Elementary simplifications complete the proof of part 5.

Part 6 will follow from part 5. If $W=0$, then $\mathbb{E}[D^4|W]=\frac{256}{n^4}$, so part 6 is valid in this case. If $W\neq 0$, then by the definition of W it follows that $W\geq \frac{4}{n}$. Note that

$$\mathbb{E}[D^4|W]-\frac{256}{n^2}W^2=\frac{64}{n^2}\left(-3W^2+\frac{6W}{n}+\frac{4}{n^2}\right)-\frac{256W^2}{n^3}-\frac{256W}{n^4}.$$

It is easy to see that $-3W^2+\frac{6W}{n}+\frac{4}{n^2}<0$ if $W\geq\frac{4}{n}$, implying that $\mathbb{E}[D^4|W]\leq\frac{256}{n^2}W^2$ if $W\neq 0$. $\square$

Proof of Proposition 4.6 One applies Theorem 1.2 to the pair (W,W'). By Part 1 of Lemma 4.7, the hypotheses are satisfied with $\lambda=\frac{4}{n}$.

Consider the first error term in Theorem 1.2. By parts 3 and 4 of Lemma 4.7,

$$\begin{aligned}\frac{\mathbb{E}|2\lambda W-\mathbb{E}[D^2|W]|}{2\lambda t} &= \frac{2}{tn}\mathbb{E}|W-1|\\ &\leq \frac{2}{tn}\sqrt{\mathbb{E}(W-1)^2}\\ &= \frac{2}{tn}.\end{aligned}$$

Consider the second error term in Theorem 1.2. By the Cauchy-Schwarz inequality,

$$\mathbb{E}|D|^3\leq\sqrt{\mathbb{E}[D^2]\mathbb{E}[D^4]}.$$

Taking expectations in part 3 Lemma 4.7 gives that $\mathbb{E}[D^2]=\frac{8}{n}$. Taking expectations in part 5 of Lemma 4.7 gives that $\mathbb{E}[D^4]=\frac{128}{n^2}-\frac{128}{n^3}\leq\frac{128}{n^2}$. Thus the second error term in Theorem 1.2 is at most $\frac{2\max\{t^{-1},2t^{-2}\}}{\sqrt{n}}$.

To bound the third error term in Theorem 1.2, one applies Theorem 3.1 with $\kappa=2$. Note from part 4 of Lemma 4.7 that $\mathbb{E}|W-1|\leq\sqrt{\mathbb{E}(W-1)^2}=1$. It is necessary to bound

$$\epsilon_1(t)=\mathbb{E}\big[\mathbb{E}(D^2|W)\mathbb{I}\{\mathbb{E}(D^2|W)>2\lambda(W+t)\text{ or }\mathbb{E}(D^4|W)>4\lambda^2(\kappa^2W^2+\kappa^2t^2)\}\big].$$

Part 3 of Lemma 4.7 implies that $\mathbb{E}[D^2|W)]>2\lambda(W+t)$ if and only if $(W-1)<-\frac{tn}{2}$. Part 5 of Lemma 4.7 implies that

$$\mathbb{E}[D^4|W)]>4\lambda^2(\kappa^2W^2+\kappa^2t^2)$$

can happen only if $W=0$. Thus

$$\epsilon_1(t)\leq\mathbb{E}\left[\mathbb{E}[D^2|W]\mathbb{I}\left\{W-1<-\frac{tn}{2}\right\}\right]\;+\;\mathbb{P}(W=0)\mathbb{E}[D^2|W=0]. \tag{22}$$

To bound the first term in (22), note by part 3 of Lemma 4.7 that

$$\mathbb{E}[D^2|W]=2\lambda+(2\lambda-\lambda^2)(W-1).$$

Since $n\geq 2$, one has that $2\lambda-\lambda^2\geq 0$. It follows that if $W-1<0$, then $\mathbb{E}[D^2|W]\leq 2\lambda$. Hence the first term in (22) is at most $\frac{8}{n}\mathbb{P}(W-1<-\frac{tn}{2})$. By Chebyshev's inequality, this is at most $\frac{32}{n^3t^2}$. To bound the second term in (22), note that $\mathbb{P}(W=0)\leq\frac{2}{n}$. Also part 3 of Lemma 4.7 gives that $\mathbb{E}[D^2|W=0]=\frac{16}{n^2}$, so that the second term in (22) is at most $\frac{32}{n^3}$. Summarizing, we have shown that $\epsilon_1(t)\leq\frac{32}{n^3}(1+\frac{1}{t^2})$.

It is also necessary to bound

$$\epsilon_2(t)=\mathbb{P}\{\mathbb{E}(D^2|W)<2\lambda(W-\tfrac{1}{4}t)\text{ or }\mathbb{E}(D^4|W)>4\lambda^2(\kappa^2W^2+\kappa^2t^2)\}.$$

Part 3 of Lemma 4.7 gives that

$$\mathbb{E}[D^2|W] < 2\lambda W - \frac{\lambda t}{2}$$

if and only if $(W-1) > \frac{tn}{8}$. Since W has mean and variance 1, Chebyshev's inequality implies that this occurs with probability at most $\frac{64}{t^2n^2}$. By part 5 of Lemma 4.7,

$$\mathbb{E}[D^4|W)] > 4\lambda^2(\kappa^2W^2+\kappa^2t^2)$$

implies that $W=0$. Since $\mathbb{P}(W=0) \leq \frac{2}{n}$, it follows that $\epsilon_2(t) \leq \frac{2}{n} + \frac{64}{t^2n^2}$.

Summarizing, the bounds on $\mathbb{E}|W-1|, \epsilon_1(t), \epsilon_2(t)$ give that the third error term in Theorem 1.2 is at most $\frac{B \cdot \max\{1,t^{-3}\}}{\sqrt{n}}$ where B is a universal constant. Adding this to the first two error terms completes the proof. □

4.3. **Lower bound.** The purpose of this subsection is to prove the lower bound from Theorem 1.4 in the introduction.

Proposition 4.8. *There is a sequence of n's tending to infinity, and corresponding t_n's such that*

$$|\mathbb{P}(W \leq t_n) - \mathbb{P}(Z \leq t_n)| = \frac{2e^{-2}}{\sqrt{n}} + O(1/n).$$

Proof. Given n, define $i = \lceil \frac{n}{2} - \sqrt{n} \rceil$ and $t_n = \frac{(n-2i)(n-2i+2)}{2n}$. The sequence of n's will consist of even perfect squares; then $i = \lceil \frac{n}{2} - \sqrt{n} \rceil = \frac{n}{2} - \sqrt{n}$ is an integer and the ceiling function can be ignored.

Clearly

$$\mathbb{P}(Z \geq t_n) = e^{(-2-\frac{2}{\sqrt{n}})} = e^{-2}\left(1 - \frac{2}{\sqrt{n}} + O(\frac{1}{n})\right).$$

Also

$$\mathbb{P}(W \geq t_n) = \mathbb{P}(i \leq \frac{n}{2} - \sqrt{n}) = \frac{\binom{n}{\frac{n}{2}-\sqrt{n}}}{\binom{n}{\frac{n}{2}}}.$$

Note that for integral a,

$$\begin{aligned}
\frac{\binom{n}{\frac{n}{2}-a}}{\binom{n}{\frac{n}{2}}} &= \frac{1}{(1+\frac{2a}{n})}\prod_{j=1}^{a-1}\frac{(1-\frac{2j}{n})}{(1+\frac{2j}{n})} \\
&= \frac{1}{(1+\frac{2a}{n})} e^{\sum_{j=1}^{a-1}[\log(1-\frac{2j}{n}) - \log(1+\frac{2j}{n})]} \\
&= \frac{1}{(1+\frac{2a}{n})} e^{\sum_{j=1}^{a-1} -\frac{4j}{n} + O(\frac{j}{n})^3}.
\end{aligned}$$

Since $a = \sqrt{n}$, one obtains that

$$\begin{aligned}
\mathbb{P}(W \geq t_n) &= \frac{1}{(1+\frac{2}{\sqrt{n}})} e^{-2+\frac{2}{\sqrt{n}}+O(\frac{1}{n})} \\
&= e^{-2} + O(1/n),
\end{aligned}$$

and the result follows. □

Remark: Similar ideas give another proof of an $O(n^{-1/2})$ upper bound for $|\mathbb{P}(W \leq t) - \mathbb{P}(Z \leq t)|$, when t is fixed. This argument was sketched to us by a referee of a much earlier (2006) version of this paper, and goes as follows. The first step is to consider $t = \frac{2(j^2+j)}{n}$ where j is integral. Then $\mathbb{P}(W \geq t) = \frac{\binom{n}{\frac{n}{2}-j}}{\binom{n}{\frac{n}{2}}}$. From page 1077 of [O], one has the asymptotics

$$\frac{\binom{n}{\frac{n}{2}-j}}{\binom{n}{\frac{n}{2}}} = e^{-\frac{2j^2}{n} + O(\frac{|j|^3}{n^2})} \tag{23}$$

for $j \leq n/4$. Since t is fixed, one has that $j = O(n^{1/2})$ and so

$$|\mathbb{P}(W \geq t) - \mathbb{P}(Z \geq t)| = |e^{-\frac{2j^2}{n} + O(\frac{|j|^3}{n^2})} - e^{\frac{-2(j^2+j)}{n}}| = O(n^{-1/2}). \tag{24}$$

The second step is to give a discretization argument allowing one to also use non-integral j. The point is that for fixed t and n growing, one can find an integer j such that $\frac{2(j^2+j)}{n} \leq t \leq \frac{2[(j+1)^2+(j+1)]}{n}$. For $j = O(n^{1/2})$, one easily checks that

$$\left| \mathbb{P}\left(Z \geq \frac{2(j^2+j)}{n}\right) - \mathbb{P}\left(Z \geq \frac{2[(j+1)^2+(j+1)]}{n}\right) \right| = O(n^{-1/2}) \tag{25}$$

and (using (23)) that

$$\left| \mathbb{P}\left(W \geq \frac{2(j^2+j)}{n}\right) - \mathbb{P}\left(W \geq \frac{2[(j+1)^2+(j+1)]}{n}\right) \right| = \frac{\binom{n}{n/2-j}}{\binom{n}{n/2}} - \frac{\binom{n}{n/2-j-1}}{\binom{n}{n/2}} = O(n^{-1/2}). \tag{26}$$

The $O(n^{-1/2})$ upper bound for $|\mathbb{P}(W \leq t) - \mathbb{P}(Z \leq t)|$ with arbitrary $t > 0$ fixed follows from (24), (25), and (26).

Appendix A. Exchangeable Pair and Moment Computations: Algebraic Approach

The purpose of this appendix is to explain an algebraic approach to the construction of the exchangeable pair (W, W') in Subsection 4.2 and to the moment computations in Lemma 4.7. Since the exchangeable pair in Subsection 4.1 is related to that of Subsection 4.2 (see the discussion in Subsection 4.2), this appendix gives insight into that exchangeable pair too. Throughout we give results for the Johnson graph $J(n,k)$, as this contains the Bernoulli-Laplace Markov chain as a special case $k = \frac{n}{2}$.

Let G be a finite group and K a subgroup of G. One calls (G,K) a Gelfand pair if the induced representation 1_K^G is multiplicity free. For background on this concept, see Chapter 3 of [D], Chapter 7 of [Mc], or Chapters 19 and 20 of [T1].

Suppose that (G,K) is a Gelfand pair, so that 1_K^G decomposes as $\bigoplus_{i=0}^s V_i$, where V_0 is the trivial module. Letting d_i be the dimension of V_i, one can define a probability measure π on $\{0, \cdots, s\}$ by $\pi(i) = \frac{d_i}{|G/K|}$. Associated to each value of i between 0 and s is a "spherical function" ω_i, which is a certain map from the double cosets of K in G to the complex numbers. Hence π can be viewed as a probability measure on spherical functions.

The spectrum of the Johnson graph $J(n,k)$ can be understood in the language of spherical functions of Gelfand pairs; this goes back to [DS], which used this viewpoint to study the convergence rate of random walk on $J(n,k)$. To describe this, suppose without loss of generality that $0 \leq k \leq \frac{n}{2}$.

Let G be the symmetric group S_n, and K the subgroup $S_k \times S_{n-k}$. Then the space G/K is in bijection with the vertices of the Johnson graph. There are $k+1$ spherical functions $\{\omega_0, \cdots, \omega_k\}$, and the dimension d_i is equal to $\binom{n}{i} - \binom{n}{i-1}$ if $1 \leq i \leq k$ and to 1 if $i = 0$. The double cosets $K_0, K_1, \cdots, K_k$ of K in G are also indexed by the numbers $0, 1, \cdots, k$; the double coset corresponding to j consists of those permutations τ in S_n such that $|\{1, \cdots, k\} \cap \{\tau(1), \cdots, \tau(k)\}| = k - j$. Letting $\omega_i(j)$ denote the value of ω_i on the double coset indexed by j, it is known that

$$\omega_i(j) = \sum_{m=0}^{i} \frac{(-i)_m (i-n-1)_m (-j)_m}{(k-n)_m (-k)_m m!}$$

where $(j)_m = j(j+1)\cdots(j+m-1)$ for $m \geq 1$ and $(j)_0 = 1$. The spectrum of random walk on the Johnson graph consists of the numbers $\omega_i(1)$ with multiplicity d_i.

Specializing to $k = \frac{n}{2}$ in the previous paragraph, the random variable W studied in Section 4 is equal to $W(i) = \frac{n}{2}\omega_i(1) + 1$, so up to constants is a random spherical function of the Gelfand pair (G, K). Section 4 of the paper [F4] used Stein's method to study random spherical functions of Gelfand pairs. Although the examples studied there were all for normal approximation, many of the results are general. For example, an exchangeable pair (W, W') was constructed using a reversible Markov chain. Specializing to the Gelfand pair corresponding to $J(n, k)$, the Markov chain is on the set $\{0, 1, \cdots, k\}$ and transitions from i to j with probability

$$L(i,j) := \frac{d_j}{|G|} \sum_{r=0}^{k} |K_r| \omega_i(K_r) \omega_1(K_r) \overline{\omega_j(K_r)}.$$

Proposition A.1 proves that the Markov chain L is a birth-death chain (and specializes to the birth-death chain of Subsection 4.2 when $k = \frac{n}{2}$). This is interesting, since from the definition of L it is not even evident that it is a birth-death chain.

Proposition A.1. *The Markov chain L on the set $\{0, 1, \cdots, k\}$ is a birth-death chain with transition probabilities*

$$L(i, i+1) = \frac{n(n+1-i)(n-i-k)(k-i)}{k(n-k)(n+1-2i)(n-2i)}$$

$$L(i, i-1) = \frac{in(n+1-i-k)(k+1-i)}{k(n-k)(n+2-2i)(n+1-2i)}$$

$$L(i, i) = \frac{i(n+1-i)(n-2k)^2}{k(n-k)(n-2i)(n+2-2i)}$$

Proof. The spherical function $\omega_i(K_r)$ is the Hahn polynomial $Q_n(x; \alpha, \beta, N)$ where $x = r, n = i, N = k, \alpha = k - n - 1, \beta = -k - 1$. Properties of these polynomials are given on pages 33-34 of [KoSw]. In particular, they satisfy a recurrence relation

$$-r\omega_i(K_r) = A_i \omega_{i+1}(K_r) - (A_i + B_i)\omega_i(K_r) + B_i \omega_{i-1}(K_r)$$

where

$$A_i = \frac{(n+1-i)(n-k-i)(k-i)}{(n+1-2i)(n-2i)}$$

and

$$B_i = \frac{i(n+1-k-i)(k+1-i)}{(n+2-2i)(n+1-2i)}.$$

Since $\omega_1(K_r) = 1 - \frac{nr}{k(n-k)}$, it follows that

$$\begin{aligned}\omega_1(K_r)\omega_i(K_r) &= \frac{n(n+1-i)(n-i-k)(k-i)}{k(n-k)(n+1-2i)(n-2i)}\omega_{i+1}(K_r) \\ &+ \frac{i(n+1-i)(n-2k)^2}{k(n-k)(n-2i)(n+2-2i)}\omega_i(K_r) \\ &+ \frac{in(n+1-i-k)(k+1-i)}{k(n-k)(n+2-2i)(n+1-2i)}\omega_{i-1}(K_r).\end{aligned}$$

The result now follows immediately from the orthogonality relations for Hahn-polynomials [KoSw], which are a special case of the orthogonality relations for spherical functions of a Gelfand pair [Mc]. □

To conclude, we note that there is an algebraic way to compute the moments $\mathbb{E}(W' - W)^m$ and the conditional moments $\mathbb{E}[(W' - W)^m|i]$. The interesting point about this approach is that it does not require one to explicitly compute the transition probabilities of the Markov chain L, or even to know that in this particular case it is a birth-death chain. Moreover, some of the quantities which appear have direct interpretations in terms of random walk on the Johnson graph.

To be precise, Lemma 4.12 of [F4] implies that $\mathbb{E}(W' - W)^m$ is equal to

$$\left(\frac{|K_1|}{|K|}\right)^{m/2} \sum_{l=0}^{m} (-1)^{m-l} \binom{m}{l} \sum_{r=0}^{s} \frac{|K|}{|K_r|} \omega_1(K_r) p_l(K_r) p_{m-l}(K_r).$$

Here $p_j(K_r)$ is the chance that random walk on the Johnson graph $J(n,k)$ started at a particular vertex, is distance r away from the start vertex after j steps. Also, the proof of the lemma gives that $\mathbb{E}[(W' - W)^m|i]$ is equal to

$$\left(\frac{|K_1|}{|K|}\right)^{m/2} \sum_{l=0}^{m} (-1)^{m-l} \binom{m}{l} \omega_i(K_1)^{m-l} \sum_{r=0}^{s} \omega_i(K_r)\omega_1(K_r)p_l(K_r).$$

These expressions are easily evaluated for small m, and one obtains another proof of Lemma 4.7.

Acknowledgements

Fulman was supported by NSA grant H98230-05-1-0031 and NSF grants DMS-0503901 and DMS-0802082.

References

[CS] Chen, L. and Shao, Q., Stein's method for normal approximation, in *An introduction to Stein's method*, Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore, Volume 4, 2005, 1-59.

[D] Diaconis, P., *Group representations in probability and statistics*, Institute of Mathematical Statistics Lecture Notes, Volume 11, 1988.

[DS] Diaconis, P. and Shahshahani, M., Time to reach stationarity in the Bernoulli-Laplace diffusion model, *Siam J. of Math. Anal.* **18** (1987), 208-218.

[F1] Fulman, J., Stein's method and Plancherel measure of the symmetric group, *Transac. Amer. Math. Soc.* **357** (2005), 555-570.

[F2] Fulman, J., Martingales and character ratios, *Trans. Amer. Math. Soc.* **358** (2006), 4533-4552.

[F3] Fulman, J., An inductive proof of the Berry-Esséen theorem for character ratios, *Ann. Comb.* **10** (2006), 319-332.

[F4] Fulman, J., Stein's method and random character ratios, *Trans. Amer. Math. Soc.* **360** (2008), 3687-3730.

[GT] Götze, F. and Tikhomirov, A. N., Limit theorems for spectra of random matrices with martingale structure, *Theory Probab. Appl.* **51** (2007), 42-64.

[Ho1] Hora, A., Central limit theorems and asymptotic spectral analysis on large graphs, *Inf. Dim. Anal. Quant. Prob. and Rel. Topics* **1** (1998), 221-246.

[Ho2] Hora, A., Central limit theorem for the adjacency operators on the infinite symmetric group, *Comm. Math. Phys.* **195** (1998), 405-416.

[IO] Ivanov, V. and Olshanski, G., Kerov's central limit theorem for the Plancherel measure on Young diagrams, in *Symmetric functions 2001: surveys of developments and perspectives*, 93-151, NATO Sci. Ser. II Math. Phys. Chem., 74, Kluwer Acad. Publ., Dordrecht, 2002.

[Ke] Kerov, S., Gaussian limit for the Plancherel measure of the symmetric group, *Compt. Rend. Acad. Sci. Paris, Serie I*, **316** (1993), 303-308.

[KoSw] Koekoek, R. and Swarttouw, R., The Askey-scheme of hypergeometric orthogonal polynomials and its q-analog, arXiv:math.CA/9602214 (1996).

[Lu] Luk, H.M., Stein's method for the gamma distribution and related statistical applications, Ph.D. thesis, University of Southern California, 1994.

[Mc] Macdonald, I., *Symmetric functions and Hall polynomials*, Second edition, Oxford University Press, New York, 1995.

[Mn] Mann, B., Stein's method for χ^2 of a multinomial, unpublished manuscript (1997).

[O] Odlyzko, A., Asymptotic enumeration methods, in *Handbook of combinatorics, Vol. 2*, 1063-1229, Elsevier, Amsterdam, 1995.

[Re] Reinert, G., Three general approaches to Stein's method, in *An introduction to Stein's method*, Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore, Volume 4 (1994), 183-221.

[RR] Rinott, Y. and Rotar, V., On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.* **7** (1997), 1080-1105.

[Ro] Röllin, A., A note on the exchangeability condition in Stein's method, *Statist. Probab. Lett.*, in press (2008).

[Sc] Scarabotti, F., Time to reach stationarity in the Bernoulli-Laplace diffusion model with many urns, *Adv. Appl. Math.* **18** (1997), 351-371.

[ShSu] Shao, Q., and Su, Z., The Berry-Esséen bound for character ratios, *Proc. Amer. Math. Soc.* **134** (2006), 2153-2159.

[Sn] Sniady, P., Gaussian fluctuations of characters of symmetric groups and of Young diagrams, *Probab. Theory Related Fields* **136** (2006), 263-297.

[St1] Stein, C., *Approximate computation of expectations*, Institute of Mathematical Statistics Lecture Notes, Volume 7, 1986.

[St2] Stein, C., with Diaconis, P., Holmes, S., and Reinert, G., Use of exchangeable pairs in the analysis of simulations, in *Stein's method: expository lectures and applications*, 69-77, IMS Lecture Notes Monogr. Ser., Volume 46, 2004.

[T1] Terras, A., *Fourier analysis on finite groups and applications*, London Mathematical Society Student Texts 43, Cambridge University Press, Cambridge, 1999.

[T2] Terras, A., Survey of spectra of Laplacians on finite symmetric spaces, *Experimental Math.*, **5** (1996), 15-32.

Department of Statistics, University of California, Berkeley, Berkeley, CA 94720-3860
E-mail address: sourav@stat.berkeley.edu

Department of Mathematics, University of Southern California, Los Angeles, CA 90089
E-mail address: fulman@usc.edu

Department of Mathematics, National University of Singapore, Singapore 117543
E-mail address: matar@nus.edu.sg