STEIN'S METHOD, MANY INTERACTING WORLDS AND QUANTUM MECHANICS

Ian W. McKeague, Erol Peköz, and Yvik Swan

Columbia University, Boston University, and Université de Liège

Abstract

Hall, Deckert and Wiseman (2014) recently proposed that quantum theory can be understood as the continuum limit of a deterministic theory in which there is a large, but finite, number of classical "worlds." A resulting Gaussian limit theorem for particle positions in the ground state, agreeing with quantum theory, was conjectured in Hall, Deckert and Wiseman (2014) and proven by McKeague and Levin (2016) using Stein's method. In this article we propose new connections between Stein's method and Many Interacting Worlds (MIW) theory. In particular, we show that quantum position probability densities for higher energy levels beyond the ground state arise as distributional fixed points in a new generalization of Stein's method. These are then used to obtain a rate of distributional convergence for conjectured particle positions in the first energy level above the ground state to the (two-sided) Maxwell distribution; new techniques must be developed for this setting where the usual "density approach" Stein solution (see Chatterjee and Shao (2011)) has a singularity.

1 INTRODUCTION

Hall, Deckert and Wiseman (2014) proposed a many interacting worlds (MIW) theory for interpreting quantum mechanics in terms of a large but finite number of classical "worlds." In the case of the MIW harmonic oscillator, an energy minimization argument was used to derive a recursion giving the location of the oscillating particle as viewed in each of the worlds. Hall et al. conjectured that the empirical distribution of these locations converges to Gaussian as the total number of worlds N increases. McKeague and Levin (2016) recently proved such a result and provided a rate of convergence. More specifically, McKeague and Levin showed that if $x_1, \dots x_N$ is a decreasing, zero-mean sequence of real numbers satisfying the recursion relation

$$x_{n+1} = x_n - \frac{1}{x_1 + \cdots + x_n}, \tag{1.1}$$

then the empirical distribution of the x_n tends to standard Gaussian when $N \to \infty$. Here x_n represents the location of the oscillating particle in the nth world, and the Gaussian limit distribution agrees with quantum theory for a particle in the lowest energy (ground) state.

The energy minimizing principle only applies to the ground state, but the hypothesized correspondence with quantum theory suggests that stable configurations should also exist at higher energies in the MIW theory. Moreover, the empirical distributions of these configurations should converge to distributions with densities of the form

$$p_k(x) = \frac{(H_k(x))^2}{k!}\varphi(x), \quad x \in \mathbb{R}, \tag{1.2}$$

where $\varphi(x)$ is the standard normal density,

$$\mathrm{He}_k(x) = (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}$$

is the (probabilist's) kth Hermite polynomial, and k is a non-negative integer. The ground state discussed above corresponds to $k = 0$ and has the standard Gaussian limit. However, the question of how to characterize higher energy MIW states corresponding to $k \geq 1$ is still unresolved as far as we know.

The energy minimization approach of Hall, Deckert and Wiseman (2014) starts with an analysis of the Hamiltonian for the MIW harmonic oscillator:

$$H(\mathbf{x}, \mathbf{p}) = E(\mathbf{p}) + V(\mathbf{x}) + U(\mathbf{x}),$$

where the locations of particles (having unit mass) in the N worlds are specified by $\mathbf{x} = (x_1, \ldots, x_N)$ with $x_1 > x_2 > \ldots > x_N$, and their momenta by $\mathbf{p} = (p_1, \ldots, p_N)$. Here $E(\mathbf{p}) = \sum_{n=1}^N p_n^2/2$ is the kinetic energy, $V(\mathbf{x}) = \sum_{n=1}^N x_n^2$ is the potential energy (for the parabolic trap), and

$$U(\mathbf{x}) = \sum_{n=1}^{N} \left(\frac{1}{x_{n+1} - x_n} - \frac{1}{x_n - x_{n-1}} \right)^2$$

is the hypothesized "interworld" potential, where $x_0 = \infty$ and $x_{N+1} = -\infty$.

In the ground state, there is no movement because all the momenta p_n have to vanish for the total energy to be minimized. In this case, as mentioned above, Hall, Deckert and Wiseman (2014) showed that the particle locations x_n satisfy (1.1) and McKeague and Levin (2016) showed that the empirical distribution tends to a standard Gaussian distribution. However, when the total energy is greater than in the ground state, the particles will move in a complex manner governed by Hamilton's equations, and a study of the time-dependent behavior of the system would require numerically solving these equations. ODE solvers that preserve the total energy of a Hamiltonian system (known as symplectic integrators) are readily available, but the actual solutions are analytically intractable. Therefore, the problem of isolating stationary solutions of the MIW harmonic oscillator to serve as parallels to higher-energy states (eigenstates) of the quantum harmonic oscillator appears to be analytically intractable.

Our proposed way around this impasse is to formulate a definition of MIW higher energy states based on a distributional fixed point property that has an analogous interpretation to the eigenstates of quantum theory. For the ground state, our approach leads to the same result as the existing MIW solution (1.1). We will focus on the second stationary state (the case $k = 1$), and obtain a recursion for the particle position that is in asymptotic agreement with the corresponding eigenstate of quantum theory. More specifically, we show that the fixed point property in this case leads to the recursion

$$x_{n+1}^3 = x_n^3 - 3N(N-1)^{-1} \left(\sum_{i=1}^{n} \frac{1}{x_i} \right)^{-1}, \tag{1.3}$$

and that if $x_1, \ldots, x_N$ is a decreasing, zero-mean solution, then the empirical distribution of the x_n converges to the (two-sided) Maxwell distribution having density $p_1(x) = x^2 e^{-x^2/2}/\sqrt{2\pi}$. We also give a rate of convergence using a new extension of Stein's method. Our approach is generalizable to recursions that converge to the distributions of other higher energy states of the quantum harmonic oscillator, although we do not pursue such extensions here.

Stein's method (see Stein (1986), Chen, Goldstein and Shao (2010) and Ross (2011)) is a well established technique for obtaining explicit error bounds for distributional limit theorems. The usual "density approach" (see Chatterjee and Shao (2011)) for applying Stein's method to arbitrary random variables does not seem to apply in cases where the density function vanishes at a point (here we have $p_1(0) = 0$); in this case the solution to the Stein equation will have a singularity and unbounded smoothness, and this therefore requires the new technique we give here to handle such distributions.

If MIW theory is to provide a satisfactory interpretation of quantum mechanics, it needs to "explain" how measurements arise from particular states of the system. In the classical Copenhagen interpretation of quantum mechanics, when position is measured the wave function is viewed as being projected onto a random eigenstate ("wave function collapse"). Eigenstates for position are fixed points of the Hamiltonian (operator) after normalization by the corresponding eigenvalue, which is in parallel to our idea of a MIW distributional fixed point, as introduced below. It is therefore reasonable to expect that a MIW measurement consisting of a random draw from the empirical distribution of $x_1, \ldots, x_N$ should agree in the limit with the distribution of the corresponding eigenstate. In the sequel we refer to such sequences $x_1, \ldots, x_N$ as *canonical configurations*.

To formalize the idea of a canonical configuration, in Section 2 we introduce the notion of a generalized zero-bias transformation, and show that the distributional properties of eigenstates of the quantum harmonic oscillator can be characterized in terms of fixed points of this transformation. Next, we derive the generalized zero-bias distribution for an empirical distribution on general configurations. In Section 3 we introduce the idea of a canonical configuration by stipulating that a smoothed version of its empirical distribution satisfies a fixed point property for the corresponding generalized zero-bias transformation. Section 4 develops general methods aimed at showing convergence of empirical distributions on canonical configurations based on the new extension of Stein's method.

2 GENERALIZED ZERO-BIAS TRANSFORMATIONS

Let W be a symmetric random variable and $b\colon \mathbb{R} \to \mathbb{R}$ a non-negative function such that $\sigma^2 = \mathbb{E}[W^2/b(W)] < \infty$. Goldstein and Reinert (1997) gives a distributional fixed point characterization of the Gaussian distribution, which we generalize in the definition below.

Definition 2.1. If there is a random variable $W^\star$ such that

$$\sigma^2 \mathbb{E}\left[\frac{f'(W^\star)}{b(W^\star)}\right] = \mathbb{E}\left[\frac{Wf(W)}{b(W)}\right]$$

for all absolutely continuous functions $f\colon \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}|Wf(W)/b(W)| < \infty$, we say that $W^\star$ has the *b-generalized-zero-bias distribution* of W.

Remark 2.2. Goldstein and Reinert (1997) study the case $b(x) = 1$ and show that $W^\star$ has the same distribution as W if and only if W has a Gaussian distribution. Distributional fixed point characterizations for exponential, gamma and other non-negative distributions and the connection with Stein's method have been studied in Peköz and Röllin (2011), Peköz, Röllin and Ross (2013a), and Peköz, Röllin and Ross (2015).

Remark 2.3. By a routine extension of the proof of Proposition 2.1 of Chen, Goldstein and Shao (2010), it can be shown that there exists a unique distribution for $W^\star$, and it is absolutely continuous with density

$$p^\star(x) = \frac{b(x)}{\sigma^2}\mathbb{E}\left[\frac{W}{b(W)}1_{W\geqslant x}\right].$$

We note in passing that the σ^2 is misplaced in the first display of Chen et al.'s proposition, which corresponds to $b(x) = 1$, the usual zero-bias distribution of W. The composition of the b-generalized-zero-bias transformation with the $(1/b)$-generalized-zero-bias transformation is the usual zero-bias transformation.

Remark 2.4. With b a φ-integrable function, if W has density

$$p(x) = b(x)\varphi(x), \tag{2.1}$$

then its distribution is a fixed point for the b-generalized-zero-bias transformation, since

$$\sigma^2 = \int_{-\infty}^{\infty} \frac{t^2}{b(t)} p(t)\,dt = \int_{-\infty}^{\infty} t^2\varphi(t)\,dt = 1$$

and

$$p^\star(x) = b(x)\int_x^\infty \frac{t}{b(t)}p(t)\,dt = b(x)\int_x^\infty t\varphi(t)\,dt = p(x).$$

The following result gives the b-generalized-zero-bias distribution of the uniform distribution on N points.

Proposition 2.5. *Given an integer $N > 1$, let $x_1 > x_2 > \ldots > x_N$ be such that $b(x_n) > 0$ for all n. Let $\mathbb{P}_N$ be the empirical distribution of the x_n:*

$$\mathbb{P}_N(A) = \frac{\#\{n\colon x_n \in A\}}{N}$$

for any Borel set $A \subset \mathbb{R}$. Under the symmetry condition $x_n = x_{N-n+1}$ for $n = 1, \ldots, N$, the b-generalized-zero-bias distribution $\mathbb{P}_N^\star$ of $\mathbb{P}_N$ is defined, and has density

$$p^\star(x) = b(x)\left[\sum_{i=1}^{n}\frac{x_i}{b(x_i)}\right]\left[\sum_{j=1}^{N}\frac{x_j^2}{b(x_j)}\right]^{-1}$$

for $x_{n+1} < x \le x_n$ $(n = 1, \ldots, N-1)$, and $p^\star(x) = 0$ if $x > x_1$ or $x \le x_N$.

Proof. Immediate from the definition of the b-generalized-zero-bias density. $\square$

Recall the following distances between distribution functions F and G. The Kolmogorov distance is

$$d_{\mathrm{K}}(F,G) = \sup_{x\in\mathbb{R}} |F(x) - G(x)|,$$

and the Wasserstein distance is

$$d_{\mathrm{W}}(F,G) = \sup_{h\in\mathcal{H}} \left| \int_{\mathbb{R}} h\, dF - \int_{\mathbb{R}} h\, dG \right|$$

where

$$\mathcal{H} = \{h \colon \mathbb{R} \to \mathbb{R} \text{ Lipschitz with } \|h'\| \leqslant 1\}$$

and $\|\cdot\|$ is the supremum norm. By assembling results from Gibbs and Su (2002), these two metrics are seen to be related by

$$d_{\mathrm{K}}(F,G) \leqslant 1.74\sqrt{d_{\mathrm{W}}(F,G)}.$$

Restricting attention to the special case $b(x) = x^2$, we can now state our main result, along with an important corollary.

Theorem 2.6. *Suppose $W^\star$ is constructed on the same probability space as the zero-mean random variable W and is distributed according to the x^2-generalized-zero-bias distribution of W. Let Z have the two-sided Maxwell density $x^2 e^{-x^2/2}/\sqrt{2\pi}$. Then there exist positive finite constants c_1, c_2, c_3 and c_4 such that*

$$d_{\mathrm{W}}(\mathscr{L}(W), \mathscr{L}(Z)) \leq c_1 \mathbb{E}\,|W - W^\star| + c_2 \mathbb{E}\,[|W|\,|W - W^\star|] \\ + c_3 \mathbb{E}\left|\frac{1}{W} - \frac{1}{W^\star}\right| + c_4 \mathbb{E}\left|1 - \frac{W^\star}{W}\right|. \tag{2.2}$$

Proof. The result follows immediately from Theorem 4.3 below. $\square$

The following corollary gives a rate of convergence of the solution to (1.3) to the two-sided Maxwell distribution in terms of the Wasserstein distance; we postpone the proof until Section 4.3.

Corollary 2.7. *Suppose $x_1, \ldots x_N$ is a monotonic, zero-mean, finite sequence of real numbers satisfying* (1.3)*, let $\mathbb{P}_N$ be the empirical distribution of these values, and let Z be as in Theorem 2.6. Then there is a constant $c > 0$ such that*

$$d_{\mathrm{W}}(\mathbb{P}_N, \mathscr{L}(Z)) \leq c\sqrt{\frac{\log N}{N}}.$$

3 CANONICAL CONFIGURATIONS

As discussed in the Introduction, we will formulate a suitable definition of a canonical configuration $\{x_n, n = 1, \ldots, N\}$ based on a fixed point property. Our proposed approach is to adapt the b-generalized-zero-bias transformation to this purpose. Note from Proposition 2.5 that $\mathbb{P}_N$ (being a discrete distribution) can never be a fixed point of the b-generalized-zero-bias transformation (since $\mathbb{P}_N^\star$ is absolutely continuous). Nevertheless, by mapping $\mathbb{P}_N$ and $\mathbb{P}_N^\star$ to slightly smoother forms we can achieve a fixed point definition as follows. Consider the following smoothed version $\tilde{\mathbb{P}}_N$ of $\mathbb{P}_N$ defined by the piecewise-constant density

$$\tilde{p}(x) = [(N-1)(x_n - x_{n+1})]^{-1}$$

for $x_{n+1} < x \le x_n$. We will refer to $\tilde{\mathbb{P}}_N$ as the *local-smooth* of $\mathbb{P}_N$. Similarly, define the local-smooth $\tilde{\mathbb{P}}_N^\star$ of the b-generalized-zero-bias distribution $\mathbb{P}_N^\star$ by the piecewise-constant density

$$\tilde{p^\star}(x) = \tilde{b}(x)\left[\sum_{i=1}^{n} \frac{x_i}{b(x_i)}\right]\left[\sum_{j=1}^{N} \frac{x_j^2}{b(x_j)}\right]^{-1},$$

where

$$\tilde{b}(x) = \int_{x_{n+1}}^{x_n} b(u)\,du \Big[x_n - x_{n+1}\Big]^{-1}$$

for $x_{n+1} < x \le x_n$. Here $\tilde{b}$ is the approximation to b that minimizes the L_2-norm under the restriction that its integral over each interval (between successive x_n) agrees with that of b. This leads to our definition of a canonical configuration based on the fixed point property of the b-generalized-zero-bias transformation after local smoothing.

Definition 3.1. A configuration $\{x_n, n = 1, \ldots, N\}$ with $x_1 > x_2 > \ldots > x_N$ is b-*canonical* if it is symmetric and the local-smooth of its empirical distribution $\mathbb{P}_N$ coincides with the local-smooth of its b-generalized-zero-bias distribution, i.e. $\tilde{\mathbb{P}}_N = \tilde{\mathbb{P}}_N^\star$.

We surmise that in MIW quantum theory, such canonical configurations play an analogous role to the eigenstates of standard quantum theory. Combining the above expressions, the existence of a b-canonical configuration is now seen to reduce to showing the existence of a symmetric solution to the recursion relation

$$\int_{x_{n+1}}^{x_n} b(u)\,du = \frac{1}{N-1}\left[\sum_{j=1}^{N} \frac{x_j^2}{b(x_j)}\right]\left[\sum_{i=1}^{n} \frac{x_i}{b(x_i)}\right]^{-1} \tag{3.1}$$

for $n = 1, \ldots, N-1$.

For the ground state of the MIW harmonic oscillator ($k = 0$), we have $b(x) = 1$ and $p^\star = \tilde{p}$, and the recursion relations (3.1) takes the simple form (1.1) under the normalization condition $\sum_{j=1}^N x_j^2 = N-1$. As shown by McKeague and Levin (2016), there is a unique decreasing zero-mean solution to this recursion, which coincides with the unique ground state of the system, and that satisfies the normalization condition and is symmetric.

For the second energy state of the MIW harmonic oscillator ($k = 1$) we have $b(x) = x^2$. In this case, we can only apply the recursion when N is even; if N is odd, the median $m = x_{(N+1)/2}$ vanishes under symmetry, so $b(m) = 0$. When N is even, the median $m = (x_{N/2} + x_{(N/2)+1})/2$ also vanishes under symmetry (and the converse is also true, as we show below), but in this case $b(x_n)$ never vanishes. The recursion (3.1) simplifies to (1.3). Fig. 1 displays the density $\tilde{p}$ having mass $1/(N-1)$

uniformly distributed over the intervals between successive x_n, compared with the Maxwell density $p_1(x) = x^2\varphi(x)$. For higher-order energy states with $k \geq 2$, it is not possible to solve the recursions explicitly, but they can be solved numerically.

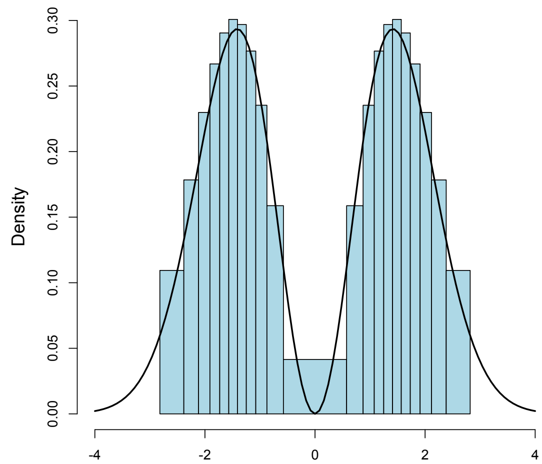

Figure 1: Example with $b(x) = x^2$, $N = 22$, showing the piecewise constant density $\tilde{p}$ of $\tilde{\mathbb{P}}_N$, the local-smooth of $\mathbb{P}_N$, compared with the Maxwell density, where the breaks in the histogram are the successive x_n satisfying the recursion (1.3).

The following lemma provides the basic properties we need to ensure the existence of a canonical configuration as a solution of the Maxwell recursion (1.3), as well as ensuring that the solution is unique. This result is analogous to Lemma 1 of McKeague and Levin (2016) concerning solutions of (1.1), but the difference here is that the variance is 3, agreeing with the Maxwell distribution (rather than close to standard normal in the case of (1.1)).

Lemma 3.2. *Suppose N is even. Every zero-median solution $x_1, \dots, x_N$ of (1.3) satisfies:*

(P1) *Zero-mean: $x_1 + \ldots + x_N = 0$.*

(P2) *Maxwell variance: $x_1^2 + \ldots + x_N^2 = 3N$.*

(P3) *Symmetry: $x_n = -x_{N+1-n}$ for $n = 1, \dots, N$.*

Further, there exists a unique solution $x_1, \dots, x_N$ such that (P1) and

(P4) *Strictly decreasing: $x_1 > \ldots > x_N$*

hold. This solution has the zero-median property, and thus also satisfies (P2) and (P3).

Proof. The proof follows identical steps to the proof of Lemma 1 of McKeague and Levin (2016), apart from the variance property (P2), which is proved using (P1) and (P3) as follows. Denote $S_n = \sum_{i=1}^n x_i^{-1}$ for $n = 1, \dots, N$, and set $S_0 = 0$. Using (1.3) we can write

$$\begin{aligned} 3N &= \frac{3N}{N-1} \sum_{n=1}^{N-1} S_n S_n^{-1} = \sum_{n=1}^{N-1} S_n (x_n^3 - x_{n+1}^3) = \sum_{n=1}^{N-1} [(S_{n-1} + x_n^{-1}) x_n^3 - S_n x_{n+1}^3] \\ &= \sum_{n=1}^{N-1} [S_{n-1} x_n^3 - S_n x_{n+1}^3 + x_n^2] \\ &= x_1^2 + \dots + x_{N-1}^2 - S_{N-1} x_N^3, \end{aligned}$$

where we used the recursion in the second equality, and the last equality is from a telescoping sum. (P3) implies $S_N = 0$, so $-S_{N-1} = 1/x_N$, and (P2) follows. $\square$

4 THE STEIN EQUATION AND ITS SOLUTIONS

4.1 General considerations

Let X have a density as in (2.1). The first step is to identify an appropriate "Stein equation" and bound its solutions. Let $\tilde{h}$ be such that $\mathbb{E}[\tilde{h}(X)] = 0$. There are many possible starting points. The well known "density approach" (see Chatterjee and Shao (2011)) starts with the Stein equation

$$f'(x) + \frac{p'(x)}{p(x)} f(x) = \tilde{h}(x)$$

which is easily solved to yield

$$f(x) = \frac{1}{b(x)\varphi(x)} \int_x^\infty \tilde{h}(u) b(u) \varphi(u)\, du.$$

If b vanishes at a point, as it does in the two-sided Maxwell case where $b(0) = 0$, then this solution f will have a singularity and we cannot carry on with the usual program for applying Stein's method. For these types of distributions we propose a new approach; the price we pay here is the necessity to bound several additional quantities concerning the couplings we obtain. The explicit nature of the recursion here allows us to compute these quantities.

In view of Definition 2.1 it is natural to consider the Stein equation

$$\frac{f'(w)}{b(w)} - w \frac{f(w)}{b(w)} = \tilde{h}(w) \tag{4.1}$$

which can be solved using the usual normal approximation solution, but the resulting estimates will again rest on properties of f/b which is unbounded in the cases we are interested in. Because of this, we introduce an original route which leads to the correct order bounds we are seeking.

First, following Ley, Reinert and Swan (2016) we introduce the integral operator associated to the Gaussian density:

$$\tilde{h} \mapsto \mathcal{T}_\varphi^{-1} \tilde{h}(w) := \frac{1}{\varphi(w)} \int_w^\infty \tilde{h}(u) \varphi(u)\, du$$

(called the "Gaussian Stein inverse operator"), which maps functions with Gaussian-mean zero and sufficiently well behaved tails into bounded functions. Next define the "Stein kernel" of X (or, equivalently, of p) by

$$\tau_X(x) = \frac{1}{p(x)} \int_x^\infty u p(u)\, du.$$

The Stein kernel satisfies the integration by parts formula

$$\mathbb{E}\left[\tau_X(X) f'(X)\right] = \mathbb{E}\left[X f(X)\right] \tag{4.2}$$

for all sufficiently regular functions f.

Remark 4.1. As mentioned in the Introduction, our concern in this paper is with symmetric densities of the form $p_k(x) = \frac{H_k^2(x)}{k!}\varphi(x) := b_k(x)\varphi(x)$. The Stein kernel of such a p_k is $\tau_k(x) = \int_x^\infty u H_k^2(u)\varphi(u)\, du / (\varphi(x) H_k^2(x)), k \geq 0$, and direct integration leads to

$$\tau_1(x) = \frac{x^2+2}{x^2},\ \tau_2(x) = \frac{x^4+2x^2+5}{(x^2-1)^2},\ \tau_3(x) = \frac{x^6+9x^2+18}{(x^3-3x)^2}. \tag{4.3}$$

The next Proposition below applies to general densities, but in the following sections we focus on further results for the special case $\tau_1(x)$.

We aim to assess the proximity between the law of X and some W by estimating the Wasserstein distance between their distributions. Suppose that there exists a $W^\star$ following the b-generalized-zero-bias distribution of W and defined on the same probability space as W. To each integrable test function h we associate the function $f := f_h$, the solution to (4.1) with $\tilde{h} = h - \mathbb{E}h(X)$. This association is unique in the sense that there exists only one absolutely continuous version of f satisfying (4.1) at all points x, see e.g. Chen, Goldstein and Shao (2010). We then write (supposing here and in the sequel that $\sigma^2 = 1$)

$$\begin{aligned} \mathbb{E}\left[h(W)\right] - \mathbb{E}\left[h(X)\right] &= \mathbb{E}\left[\frac{f'(W)}{b(W)} - W\frac{f(W)}{b(W)}\right] \\ &= \mathbb{E}\left[\frac{f'(W)}{b(W)} - \frac{f'(W^\star)}{b(W^\star)}\right]. \end{aligned} \tag{4.4}$$

As mentioned above, the function f in (4.4) is not necessarily sufficiently well-behaved for our purposes. We will therefore use the following alternative approach.

Proposition 4.2. *Let $x \mapsto b(x)$ be a nonnegative even function with support in $(-\infty, \infty)$ such that $\lim_{x\to\pm\infty} b(x)\varphi(x) = 0$. Suppose furthermore that b is absolutely continuous and integrable w.r.t. φ with integral $\int_{-\infty}^\infty b(x)\varphi(x)\, dx = 1$. Let X be a random variable with density $x \mapsto b(x)\varphi(x)$. Then*

$$\tau_X(x) = 1 + \frac{\mathcal{T}_\varphi^{-1} b'(x)}{b(x)} \tag{4.5}$$

under the convention that the ratio is set to zero at all points x such that $b(x) = 0$ and $\mathcal{T}_\varphi^{-1} b'(x) \neq 0$. Further, with $\tilde{h}$ defined by (4.1),

$$g_h(x) = \frac{\int_x^\infty b(u)\tilde{h}(u)\varphi(u)\, du}{b(x)\varphi(x) + \int_x^\infty b'(u)\varphi(u)\, du} \tag{4.6}$$

is the unique bounded solution of the ODE

$$\tau_X(x)g'(x) - xg(x) = \tilde{h}. \tag{4.7}$$

Proof. The first claim (as well the necessary conditions under which this claim holds) follows from integrating by parts in the identity

$$\int_x^{+\infty} yp(y)\,dy = \int_x^{\infty} b(y)(-\varphi'(y))\,dy.$$

To see the second claim note how any bounded solution to (4.7) must be of the form

$$g(x) = \frac{1}{\tau_X(x)\varphi(x)b(x)} \int_x^{\infty} \tilde{h}(y)b(y)\varphi(y)\,dy,$$

so that the claim follows by applying (4.5). □

Intuition (supported e.g. by Stein (1986, Lesson VI) or the more recent work Döbler (2015)) encourages us to claim that functions such as (4.6) will have satisfactory behavior. It is thus natural to seek a connection between equations of the form (4.1) and (4.7). To this end we introduce a function $g = g_f$, say, such that

$$\frac{f'(x) - xf(x)}{b(x)} = \tau_X(x)g'(x) - xg(x).$$

Since

$$\frac{f'(x) - xf(x)}{b(x)} = \frac{(f(x)\varphi(x))'}{b(x)\varphi(x)}$$

and

$$\tau_X(x)g'(x) - xg(x) = \frac{(b(x)\tau_X(x)g(x)\varphi(x))'}{b(x)\varphi(x)}$$

at all x for which $b(x) \neq 0$, we deduce that f and g are mutually defined by $f = (b\tau_X)g$. This in turn gives

$$\frac{f'(x)}{b(x)} = \left(\frac{b'(x)}{b(x)}\tau_X(x) + \tau_X'(x)\right) g(x) + \tau_X(x)g'(x) =: \psi(x)g(x) + \tau_X(x)g'(x)$$

which, combined with $\psi(x) = x(\tau_X(x) - 1)$ (that is easily derived using the various definitions involved), leads to the useful identity

$$\frac{f'(x)}{b(x)} = x(\tau_X(x) - 1)g(x) + \tau_X(x)g'(x). \tag{4.8}$$

Plugging (4.8) into (4.4), we finally obtain

$$\begin{aligned}\mathbb{E}\left[h(W)\right] - \mathbb{E}\left[h(X)\right] = \mathbb{E}\left[W(\tau_X(W) - 1)g(W) - W^\star(\tau_X(W^\star) - 1)g(W^\star)\right] \\ + \mathbb{E}\left[\tau_X(W)g'(W) - \tau_X(W^\star)g'(W^\star)\right].\end{aligned} \tag{4.9}$$

It then remains to find bounds on the two terms on the rhs of (4.9).

4.2 Approximating the two-sided Maxwell distribution

Theorem 4.3. *Let $b(x) = x^2$, $p(x) = x^2\varphi(x)$, and take f to be a solution of the Stein equation*

$$f'(w)/b(w) - wf(w)/b(w) = \tilde{h}(w), \tag{4.10}$$

where $\tilde{h}$ a function having bounded first derivative and zero-mean under p. Then there exist positive finite constants $\lambda_1, \lambda_2, \lambda_3$ and λ_4 such that

$$\begin{aligned}\left|\mathbb{E}\left[\frac{f'(W)}{(W)^2} - \frac{f'(W^\star)}{(W^\star)^2}\right]\right| &\leq \lambda_1 \mathbb{E}\left|W - W^\star\right| + \lambda_2 \mathbb{E}\left[|W|\left|W - W^\star\right|\right] \\ &\quad + \lambda_3 \mathbb{E}\left|\frac{1}{W} - \frac{1}{W^\star}\right| + \lambda_4 \mathbb{E}\left|1 - \frac{W^\star}{W}\right|. \end{aligned} \tag{4.11}$$

These constants are

$$\lambda_1 = \|\chi\|, \quad \lambda_2 = \|\chi'\|, \quad \lambda_3 = 2\left(\|g\| + \|\chi\|\right) \text{ and } \lambda_4 = 2\left(\|g'\| + \|\chi'\|\right)$$

where

$$g(x) = \frac{\int_x^\infty \tilde{h}(u)p(u)\,du}{\left(1 + \frac{2}{x^2}\right)p(x)} \text{ and } \chi(x) = \frac{g'(x)}{x}. \tag{4.12}$$

Proof. If $b(x) = x^2$ then $\tau_X(x) = 1 + 2/x^2$ and $\psi(x) = 2/x$ so that (4.9) becomes

$$\begin{aligned} &= \mathbb{E}\left[\frac{2}{W^\star}g(W^\star) - \frac{2}{W}g(W)\right] + \mathbb{E}\left[\left(1 + \frac{2}{(W^\star)^2}\right)g'(W^\star) - \left(1 + \frac{2}{(W)^2}g'(W)\right)\right] \\ &= 2\mathbb{E}\left[\left(\frac{1}{W^\star} - \frac{1}{W}\right)g(W^\star)\right] + 2\mathbb{E}\left[\frac{1}{W}\left(g(W^\star) - g(W)\right)\right] \\ &\quad + \mathbb{E}\left[g'(W^\star) - g'(W)\right] + 2\mathbb{E}\left[\frac{1}{(W^\star)^2}g'(W^\star) - \frac{1}{(W)^2}g'(W)\right]. \end{aligned}$$

The first two terms are dealt with easily to get

$$\begin{aligned} &2\left|\mathbb{E}\left[\left(\frac{1}{W^\star} - \frac{1}{W}\right)g(W^\star)\right] + 2\mathbb{E}\left[\frac{1}{W}\left(g(W^\star) - g(W)\right)\right]\right| \\ &\leq 2\|g\|\mathbb{E}\left[\left|\frac{1}{W^\star} - \frac{1}{W}\right|\right] + 2\|g'\|\mathbb{E}\left[\frac{1}{|W|}\left|W^\star - W\right|\right]. \end{aligned}$$

For the last two terms we introduce the function

$$\chi(x) = g'(x)/x$$

to get on the one hand

$$\begin{aligned} \mathbb{E}\left[g'(W^\star) - g'(W)\right] &= \mathbb{E}\left[W^\star\frac{g'(W^\star)}{W^\star} - W\frac{g'(W)}{W}\right] \\ &= \mathbb{E}\left[(W^\star - W)\chi(W^\star)\right] + \mathbb{E}\left[W\left(\chi(W^\star) - \chi(W)\right)\right] \end{aligned}$$

so that

$$\left|\mathbb{E}\left[g'(W^\star) - g'(W)\right]\right| \leq \|\chi\|E\left[|W^\star - W|\right] + \|\chi'\|E\left[|W(W^\star - W)|\right]$$

and, on the other hand

$$\mathbb{E}\left[\frac{1}{(W^\star)^2}g'(W^\star) - \frac{1}{(W)^2}g'(W)\right] = \mathbb{E}\left[\frac{1}{W^\star}\chi(W^\star) - \frac{1}{W}\chi(W)\right]$$
$$= \mathbb{E}\left[\left(\frac{1}{W^\star} - \frac{1}{W}\right)\chi(W^\star)\right] + \mathbb{E}\left[\frac{1}{W}\left(\chi(W^\star) - \chi(W)\right)\right]$$

so that

$$2\left|\mathbb{E}\left[\frac{1}{(W^\star)^2}g'(W^\star) - \frac{1}{(W)^2}g'(W)\right]\right| \le 2\|\chi\|\mathbb{E}\left[\left|\frac{1}{W^\star} - \frac{1}{W}\right|\right] + 2\|\chi'\|\mathbb{E}\left[\frac{1}{|W|}|W^\star - W|\right].$$

Combining these different estimates we obtain (4.11). The finiteness of the constants $\lambda_1, \lambda_2, \lambda_3$ and λ_4 follows from Proposition 4.4 below. □

It remains to bound $\lambda_1, \lambda_2, \lambda_3$ and λ_4 in a non trivial way. First note that if $X \sim p$ then $\mathbb{E}[h(X)] = \mathbb{E}[N^2 h(N)]$ with $N \sim \varphi$ a standard Gaussian random variable. It will be useful to rewrite g as

$$g(x) = \frac{g_0(x)}{x^2+2}, \tag{4.13}$$

with g_0 a solution to the Stein equation

$$g_0'(x) - xg_0(x) = x^2\left(h(x) - \mathbb{E}[N^2h(N)]\right). \tag{4.14}$$

Such a function g_0 exists because

$$\int_{-\infty}^{\infty} u^2\left(h(u) - \mathbb{E}[N^2h(N)]\right)\varphi(u)\,du = \mathbb{E}[N^2h(N)] - \int_{-\infty}^{\infty} u^2\varphi(u)\,du\mathbb{E}[N^2h(N)] = 0,$$

so (4.14) does indeed possess a solution. The following identity will be useful:

$$e^{x^2/2}\int_x^\infty u^2 e^{-u^2/2}\,du = x + e^{x^2/2}\int_x^\infty e^{-u^2/2}\,du \tag{4.15}$$

(this follows e.g. by integration by parts). We are now ready to provide bounds on $\|g\|, \|g'\|, \|\chi\|$ and $\|\chi'\|$.

Proposition 4.4. *Let h be integrable with respect to $x^2\varphi(x)$, let $\tilde{h} = h - \mathbb{E}N^2h(N)$ and set $c_0 = \|\tilde{h}\|$ and $c_1 = \|h'\|$ if h is absolutely continuous, $c_1 = +\infty$ otherwise. Then*

$$\lambda_1 = \|\chi\| \le 2\min(c_0, 2c_1) + \min\left(c_0, \sqrt{2/\pi}\right),$$
$$\lambda_2 = \|\chi'\| \le 4\min(c_0, c_1) + \sqrt{2/\pi} + \frac{c_1}{2},$$
$$\lambda_3 = 2(\|g\| + \|\chi\|) \le 8\min(c_0, 2c_1) + 2\min\left(c_0, \sqrt{2/\pi}\right),$$
$$\lambda_4 = 2(\|g'\| + \|\chi'\|) \le 8\min(c_0, c_1) + 2\sqrt{2/\pi} + c_1.$$

Proof. Step 1: a bound on $\|g\|$. It follows (by a similar argument as Chen, Goldstein and Shao (2010, Proof of Lemma 2.4, page 39)) that

$$|g_0(x)| \leq \begin{cases} e^{x^2/2}\int_{-\infty}^{x} u^2|\tilde{h}(u)|e^{-u^2/2}\,du & \text{if } x \leq 0 \\ e^{x^2/2}\int_{x}^{\infty} u^2|\tilde{h}(u)|e^{-u^2/2}\,du & \text{if } x \geq 0 \end{cases}$$
$$\leq e^{x^2/2}\min\left(c_0\int_{|x|}^{\infty} u^2 e^{-u^2/2}\,du, c_1\int_{|x|}^{\infty} u^2\left(|u|+2\sqrt{2/\pi}\right)e^{-u^2/2}\,du\right).$$

Using (4.15) for the first term and a similar argument for the second term we can rewrite the above as

$$|g_0(x)| \leq \min\left\{c_0\left(|x|+e^{x^2/2}\int_{|x|}^{\infty} e^{-u^2/2}\,du\right),\right.$$
$$\left. c_1\left(x^2+2\sqrt{\frac{2}{\pi}}|x|+2+e^{x^2/2}\int_{|x|}^{\infty} e^{-u^2/2}\,du\right)\right\}.$$

Applying the well-known bound

$$e^{x^2/2}\int_{|x|}^{\infty} e^{-u^2/2}\,du \leq \sqrt{\pi/2}$$

we get

$$|g_0(x)| \leq \min\left\{c_0\left(|x|+\sqrt{\frac{\pi}{2}}\right), c_1\left(x^2+2\sqrt{\frac{2}{\pi}}|x|+2+\sqrt{\frac{\pi}{2}}\right)\right\}.$$

Finally using the easily established facts

$$\frac{|x|+\sqrt{\pi/2}}{x^2+2} \leq 1 \text{ and } \frac{x^2+2\sqrt{\frac{2}{\pi}}|x|+2+\sqrt{\frac{\pi}{2}}}{x^2+2} \leq 2$$

we can conclude

$$|g(x)| = \left|\frac{g_0(x)}{x^2+2}\right| \leq \min\left(c_0, 2c_1\right) \tag{4.16}$$

and thus (we multiply the first term by 2 in order to simplify future expressions)

$$\|g\| \leq 2\min\left(c_0, c_1\right). \tag{4.17}$$

Step 2: a bound on $\|g'\|$. Taking the derivative of (4.13) we get

$$g'(x) = \frac{g_0'(x)}{x^2+2} - \frac{2x}{(x^2+2)^2}g_0(x).$$

We bound the two terms separately. For the second we use the fact that

$$\frac{|2x|}{x^2+2} \leq \frac{\sqrt{2}}{2} \leq 1$$

so that

$$\left|\frac{2x}{(x^2+2)^2}g_0(x)\right| \leq \frac{|g_0(x)|}{x^2+2}$$

and we can apply (4.16). For the first term we again follow Chen, Goldstein and Shao (2010, Proof of Lemma 2.4, page 39) to conclude that

$$\frac{|g_0'(x)|}{x^2+2} \leq \frac{x^2}{x^2+2}\left|h(x)-\mathbb{E}[N^2h(N)]\right| + c_0\frac{x}{x^2+2}xe^{x^2/2}\int_x^\infty u^2e^{-u^2/2}\,du.$$

A similar argument holds for $x \leq 0$ and we conclude

$$\left|\frac{g_0'(x)}{x^2+2}\right| \leq 2c_0. \tag{4.18}$$

If h is only absolutely continuous then we need to adapt the corresponding part of Chen, Goldstein and Shao (2010, Proof of Lemma 2.4, page 39). Following this proof we obtain the bound

$$\begin{aligned}\frac{|g_0'(x)|}{x^2+2} \leq c_1&\frac{\left(x^2-\sqrt{2\pi}xe^{x^2/2}(1-\Phi(x))\int_{-\infty}^x y^2\varphi(y)\,dy\right)}{x^2+2}\\ &+c_1\frac{\left(x^2+\sqrt{2\pi}xe^{x^2/2}\Phi(x)\int_x^\infty y^2\varphi(y)\,dy\right)}{x^2+2} \leq 2c_1\end{aligned} \tag{4.19}$$

(here we use the notation $\Phi(x) = \int_{-\infty}^x \varphi(y)\,dy$). Combining (4.18) and (4.19) we conclude

$$\left|\frac{g_0'(x)}{x^2+2}\right| \leq \min(2c_0, 2c_1). \tag{4.20}$$

Combining (4.20) and (4.16) leads to

$$\|g'\| \leq 4\min(c_0, c_1). \tag{4.21}$$

Step 3: a bound on $\chi(x) = g'(x)/x$. Using (4.14) we know that

$$\begin{aligned}g'(x) &= \frac{g_0'(x)}{x^2+2} - \frac{2x}{(x^2+2)^2}g_0(x)\\ &= \frac{xg_0(x)+x^2\left(h(x)-\mathbb{E}[N^2h(N)]\right)}{x^2+2} - \frac{2x}{(x^2+2)^2}g_0(x)\end{aligned}$$

so that

$$\chi(x) = \left(1-\frac{2}{x^2+2}\right)\frac{g_0(x)}{x^2+2} + \frac{x}{x^2+2}\left(h(x)-\mathbb{E}[N^2h(N)]\right).$$

Using

$$\left|1-\frac{2}{x^2+2}\right| \leq 1$$

as well as (4.16) we deduce

$$\|\chi\| \le \min(c_0, 2c_1) + \left\| \frac{x}{x^2+2}\left(h(x) - \mathbb{E}[N^2 h(N)]\right) \right\|.$$

Also

$$\begin{aligned}\left\| \frac{x}{x^2+2}\left(h(x) - \mathbb{E}[N^2 h(N)]\right) \right\| &\le \min\left(\frac{|x|}{x^2+2}c_0, \frac{|x|}{x^2+2}(c_1|x| + 2\sqrt{2/\pi} \right) \\ &\le \min\left(\frac{c_0}{2}, c_1 + \sqrt{2/\pi} \right) \end{aligned} \tag{4.22}$$

so that

$$\|\chi\| \le 2\min(c_0, 2c_1) + \min\left(c_0, \sqrt{2/\pi}\right). \tag{4.23}$$

Step 4: a bound on $\|\chi'\|$. Direct computations yield

$$\begin{aligned}\chi'(x) = &-\frac{2x(x^2-2)}{(x^2+2)^2}\frac{g_0(x)}{x^2+2} + \left(1 - \frac{2}{x^2+2}\right)\frac{g_0'(x)}{x^2+2} \\ &+ \frac{2-x^2}{x^2+2}\frac{1}{x^2+2}\left(h(x) - \mathbb{E}[N^2 h(N)]\right) + \frac{x}{x^2+2}h'(x).\end{aligned}$$

Using

$$\frac{|2x(x^2-2)|}{(x^2+2)^2} \le 1, \qquad \left|1 - \frac{2}{x^2+2}\right| \le 1 \text{ and } \frac{2-x^2}{x^2+2} \le 1$$

as well as (4.16) and (4.20) and a bound similar to (4.22) we get

$$\|\chi'\| \le \min(c_0, 2c_1) + 2\min(c_0, c_1) + \min\left(\frac{c_0}{2}, c_1 + \sqrt{2/\pi}\right) + \frac{c_1}{2}.$$

which we approximate (rather roughly) with

$$\|\chi'\| \le 4\min(c_0, c_1) + \sqrt{2/\pi} + \frac{c_1}{2}. \tag{4.24}$$

□

Remark 4.5. It can be seen that g'' is not bounded. Also throughout the proof we imposed no condition on the derivative of h and it is only at the very last step (in the bound (4.24)) that we restrict to smooth test functions.

4.3 Verifying the moment conditions

In this section we find bounds on the moments in Theorem 2.6 in order to prove Corollary 2.7. We will make use of the following lemma.

Lemma 4.6. *If $x_1, \dots, x_N$ is the unique strictly decreasing zero-mean solution of (1.3), then $x_1 = O(\sqrt{\log N})$.*

Proof. To simplify the notation, note that it suffices to consider the rescaled recursion $x_{n+1}^3 = x_n^3 - S_n^{-1}$, where S_n is defined in the proof of Lemma 3.2. By expressing x_1^3 as a telescoping sum,

$$x_1^3 = \sum_{n=1}^{m-1}(x_n^3 - x_{n+1}^3) + x_m^3 = \sum_{n=1}^{m-1} S_n^{-1} + x_m^3 \le \sum_{n=1}^{m-1}(n/x_1)^{-1} + x_m^3 \le x_1(1+\log m) + x_m^3,$$

where we have used Euler's approximation to the harmonic sum for the last inequality. By the variance property (P2) (in this rescaled case $x_1^2 + \ldots + x_N^2 = N-1$) we have that x_1 is bounded away from zero (as a sequence indexed by N) and x_m is bounded, so x_m^2/x_1 is bounded. Dividing the above display by x_1, we then obtain $x_1 = O(\sqrt{\log N})$. $\square$

We now prove Corollary 2.7.

Proof. (of Corollary 2.7)). From Lemma 4.6 we have

$$\mathbb{E}|W - W^\star| \le \frac{1}{N-1}\sum_{n=1}^{N-1}(x_n - x_{n+1}) = \frac{2x_1}{N-1} = O\left(\frac{\sqrt{\log N}}{N}\right). \tag{4.25}$$

Second, using $|W| \le x_1 = O(\sqrt{\log N})$ it follows immediately that

$$\mathbb{E}[|W||W - W^\star|] = O\left(\frac{\log N}{N}\right).$$

Third, the zero-median property gives

$$2x_m^3 = x_m^3 - x_{m+1}^3 = S_m^{-1} \ge (m/x_m)^{-1} = x_m/m,$$

where $m = N/2+1$, so $x_m \ge 1/\sqrt{N}$. By symmetry

$$\mathbb{E}\left|\frac{1}{W} - \frac{1}{W^\star}\right| = \mathbb{E}\left|\frac{1}{W} - \frac{1}{W^\star}\right| 1_{W^\star \in (x_{m+1}, x_m]} + 2\sum_{n=1}^{m-1}\mathbb{E}\left|\frac{1}{W} - \frac{1}{W^\star}\right| 1_{W^\star \in (x_{n+1}, x_n]}.$$

From its definition, $p^\star(x) \propto x^2$ for $x \in (x_{m+1}, x_m]$ and puts total mass $1/(N-1)$ on that interval, so the first term above can be written

$$\frac{6}{x_m^3(N-1)}\int_0^{x_m}\left(\frac{1}{x} - \frac{1}{x_m}\right)x^2\,dx \le \frac{3}{x_m(N-1)} = O\left(\frac{1}{\sqrt{N}}\right).$$

The second term is bounded above by the telescoping sum

$$\frac{2}{N-1}\sum_{n=1}^{m-1}\left(\frac{1}{x_{n+1}} - \frac{1}{x_n}\right) = \frac{2}{N-1}\left(\frac{1}{x_m} - \frac{1}{x_1}\right) = O\left(\frac{1}{\sqrt{N}}\right),$$

so we have

$$\mathbb{E}\left|\frac{1}{W} - \frac{1}{W^\star}\right| = O\left(\frac{1}{\sqrt{N}}\right).$$

Fourth,

$$\mathbb{E}\left|1 - \frac{W^\star}{W}\right| \le \sqrt{N}\,\mathbb{E}|W - W^\star| = O\left(\sqrt{\frac{\log N}{N}}\right)$$

using $|W| \ge x_m \ge 1/\sqrt{N}$ and (4.25). The Corollary follows now from Theorem 2.6. $\square$

ACKNOWLEDGEMENTS

The research of Ian McKeague was partially supported by NSF Grant DMS-1307838 and NIH Grant 2R01GM095722-05. The research of Yvik Swan was partially supported by F.R.S.-FNRS as well as IAP Research Network P7/06 of the Belgian State (Belgian Science Policy). We also thank the Institute for Mathematical Sciences at National University of Singapore for support during the *Workshop on New Directions in Stein's Method* (May 18–29, 2015) where work on the paper was initiated.

REFERENCES

S. Chatterjee and Q.-M. Shao (2011). Non-normal approximation by Stein's method of exchangeable pairs with application to the Curie–Weiss model. *Ann. App. Probab.* **21**, 464–483.

L. Chen, L. Goldstein and Q.-M. Shao (2010). *Normal Approximation by Stein's Method*. Springer Verlag.

C. Döbler (2015). Stein's method of exchangeable pairs for the beta distribution and generalizations, *Electron. J. Probab.* **20**, 1–34.

A. L. Gibbs, and F. E. Su. (2002). On choosing and bounding probability metrics. *International Statistical Review* **70**, 419–435.

L. Goldstein and G. Reinert (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7**, 935–952.

M. J. W. Hall, D. A. Deckert and H. M. Wiseman (2014). Quantum phenomena modeled by interactions between many classical worlds. *Phys. Rev. X* **4**, 041013.

C. Ley, G. Reinert and Y. Swan (2016). Stein's method for comparison of univariate distributions http://arxiv.org/abs/1408.2998

I. W. McKeague and B. Levin (2016). Convergence of empirical distributions in an interpretation of quantum mechanics. http://arxiv.org/abs/1412.1563, *Ann. Appl. Probab.* to appear.

E. Peköz and A. Röllin (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.* **39**, 587–608.

E. Peköz, A. Röllin and N. Ross (2013a). Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Probab.* **23**, 1188–1218.

E. Peköz, A. Röllin and N. Ross (2015). Generalized gamma approximation with rates for urns, walk and trees. To appear in *Annals of Probability.* Preprint at `http://arxiv.org/abs/1309.4183`.

N. Ross (2011). Fundamentals of Stein's method. *Probab. Surv.* **8**, 210–293.

C. Stein (1986). *Approximate computation of expectations.* Institute of Mathematical Statistics Lecture Notes–Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA.