Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2013, Vol. 49, No. 2, 529–549
DOI: 10.1214/11-AIHP473

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Stein's method in high dimensions with applications

Adrian Röllin

Department of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, 117546 Singapore

Received 1 February 2011; revised 19 September 2011; accepted 26 December 2011

Abstract. Let h be a three times partially differentiable function on $\mathbb{R}^n$, let $X = (X_1, \dots, X_n)$ be a collection of real-valued random variables and let $Z = (Z_1, \dots, Z_n)$ be a multivariate Gaussian vector. In this article, we develop Stein's method to give error bounds on the difference $\mathbb{E}h(X) - \mathbb{E}h(Z)$ in cases where the coordinates of X are not necessarily independent, focusing on the high dimensional case $n \to \infty$. In order to express the dependency structure we use Stein couplings, which allows for a broad range of applications, such as classic occupancy, local dependence, Curie–Weiss model, etc. We will also give applications to the Sherrington–Kirkpatrick model and last passage percolation on thin rectangles.

Résumé. Soit h une fonction réelle sur $\mathbb{R}^n$ dont les dérivées partielles d'ordre trois existent, soit $X = (X_1, \dots, X_n)$ un vecteur de variables aléatoire réelles et soit $Z = (Z_1, \dots, Z_n)$ un vecteur aléatoire Gaussien. Dans cet article, nous établissons par la méthode de Stein une majoration de la différence $\mathbb{E}h(X) - \mathbb{E}h(Z)$ dans le cas où les coordonnées de X ne sont pas nécessairement indépendantes; nous nous concentrons sur le cas de la grande dimension $n \to \infty$. Pour exprimer la structure de dépendance, nous utilisons des couplages de Stein, ce qui permet une large gamme d'applications, par exemple aux modèles d'urnes, au modèles avec dépendance locale, au modèle de Curie–Weiss, etc. Nous présentons aussi des applications au modèle de Sherrington–Kirkpatrick et à la percolation de dernier passage dans des rectangles étroits.

MSC: 60F17; 82B44

Keywords: Stein's method; Gaussian interpolation; Last passage percolation on thin rectangles; Sherrington–Kirkpatrick model; Curie–Weiss model

1. Introduction

Let X and Z be random vectors in $\mathbb{R}^n$ and let $h:\mathbb{R}^n \to \mathbb{R}$ be a function of interest. A fundamental problem in probability theory is to obtain bounds on the quantity

$$\left|\mathbb{E}h(X) - \mathbb{E}h(Z)\right|, \tag{1.1}$$

that is, to estimate the error when we replace X in $\mathbb{E}h(X)$ by Z. If the error in (1.1) is small irrespective of the detailed properties of X and Z then we will attribute to the function h a certain degree of *universality*, which means that the expected value only depends on certain basic characteristics of X and Z, such as the first few moments.

Of particular interest is the case where Z is a Gaussian vector having the same (or a similar) covariance structure as X, and probably the most prominent occurrence of such universality is the central limit theorem. If X is a random vector, such that the X_i are independent of each other, centred and scaled such that $\sum_i \operatorname{Var} X_i = 1$, and Z is a centred Gaussian vector with uncorrelated coordinates having the same variances as those of X, then it is well known that (1.1) is small for functions of the form

$$h(x) = g\left(\sum_{i=1}^n x_i\right), \tag{1.2}$$

where $g:\mathbb{R}\to\mathbb{R}$ is not too irregular. A common heuristic says that the central limit theorem will also hold if independence is replaced by some form of "weak" dependence, and, furthermore, it can be expected that in many cases (1.1) will be small for more general functions than (1.2). Thus, in terms of dropping independence and considering more general functions than (1.2), universality often can be observed beyond the standard setting of the central limit theorem.

Even if the vector X is such that $\sum_i X_i$ does not satisfy the central limit theorem, we can consider (1.1) for functions more general than (1.2). Let, for example, ξ_i be the number of balls that end up in the ith urn, when a fixed number of balls m is distributed independently among n urns. Clearly, $\sum_i \xi_i = m$, respectively, $\sum_i X_i = 0$ if the X_i are the centered and properly scaled ξ_i. Although these sums do not satisfy a central limit theorem, it is nevertheless possible to give informative bounds on (1.1). Dembo and Rinott [11] and Chen and Röllin [10] considered, for example, functions of the form $h(x)=g(\sum_i \varphi(x_i))$ for fixed functions g and φ, where φ is non-linear. For other, non-trivial choices of h we refer to Sections 4.1 and 4.2.

Over the last decades, Stein's method has proved to be a very robust method to obtain explicit bounds for univariate and multivariate distributional approximations in cases where X exhibits non-trivial dependencies which are not of martingale type, but more combinatorial in flavour. Although Stein's method for the multivariate normal distribution has been successfully implemented in many places (see Meckes [18] and Reinert and Röllin [21] and references therein), the dependence on the dimension of the results obtained so far may give the impression that the method is not suitable if the dimension grows linearly with the size of the problem. Indeed, this high-dimensional case has remained untackled until now. The purpose of this article is to close this gap.

It is important to note at this point that the type of bounds that we will obtain will generally not imply that the marginal distributions of the individual coordinates will converge to a normal distribution. That is, the aim is not to prove convergence to a multivariate normal distribution. In the already mentioned example of classic occupancy, if the number of balls and urns are of the same order, then ξ_i will converge to a Poisson distribution with mean being equal to the limiting ratio $\lim m/n$. Bounds on (1.1) will only be informative if they are smaller than the fluctuation of $h(X)$, that is, if the bounds are smaller than $\mathbb{E}|h(X)|$ (assuming here without loss of generality that $\mathbb{E}h(Z)=0$), which is an obvious upper bound on (1.1). The bounds that we obtain for functions h that concentrate only on a few coordinates will typically have the same order as $\mathbb{E}|h(X)|$ and hence will not – and often cannot – be informative.

The remainder of the article is organised as follows. In the next section we will first discuss the key tools used in this article, in particular the fundamental idea of using interpolation to estimate (1.1), the Gaussian integration by parts formula and multivariate Stein couplings, leading to our main result, Lemma 2.1. In Section 3 we will then give some abstract and more concrete examples of Stein couplings, ranging from the independent case to more sophisticated dependencies. In Section 4 we will discuss various applications.

2. The key lemma

An old idea to compare two quantities of interest is to find an interpolating sequence between them and to estimate the error "along the way" of the interpolation using the derivatives of h (paraphrasing Talagrand [29] on "Gaussian interpolation and the smart path method"). One of the earliest encounters of this idea is Lindeberg's method of telescoping sums. Define the interpolating sequence

$$Y(i)=(X_1,\dots,X_i,Z_{i+1},\dots,Z_n), \tag{2.1}$$

and write

$$\mathbb{E}h(X)-\mathbb{E}h(Z)=\sum_{i=1}^{n}\mathbb{E}\bigl\{h\bigl(Y(i)\bigr)-h\bigl(Y(i-1)\bigr)\bigr\}; \tag{2.2}$$

one can now bound the right-hand side of (2.2) using Taylor expansion; this idea has been successfully implemented by Rotar' [24], Chatterjee [7], Mossel et al. [19] and Tao and Vu [30] and surely by other authors. One of the important consequences of this approach is apparent when we look at (2.1): it forces us to treat the coordinates of X in an ordered way. If the components of X are independent or, more generally, a martingale difference sequence, then this is of course desirable, and, indeed, quite a few central limit theorems for martingales are based upon (2.2) (see e.g.

Bolthausen [4], Grama [15] and Rinott and Rotar' [22]). And even if no such structure is apparent in the problem, one can sometimes arrange X such that it will be close enough to a martingale difference sequence.

This approach, however, is not entirely satisfying. Often the martingale structure is "artificial" and one would like to make use of a more natural dependence structure in X, instead (rates of convergences being another reason to avoid martingales). And in some cases, one may have difficulties to linearise the problem at all.

A key difference in Stein's method is to chose an interpolating sequence that, in contrast to Lindeberg's telescoping sum, treats the components of X *symmetrically*. Note that (2.1) essentially interpolates "along the coordinate axes" and the order of the axes determines the linearisation of the problem. Instead, we will interpolate between X and Z in a way that will linearly interpolate between the matrices XX^t and ZZ^t. This approach is well-known as *Gaussian interpolation* and independently developed by Slepian [25] and Stein [26], although the technique used by Stein looks very much different from what is commonly referred to as Gaussian interpolation (the interpolation is "hidden" in the solution to the so-called *Stein equation*).

Gaussian interpolation has become popular in many areas; Talagrand [29] gives a good account of the key idea, in particular in the context of statistical mechanics (where Gaussian interpolation is referred to as *smart path method*). The method is a key ingredient in the rigorous proof of the *Parisi formula* by Talagrand [28]. It is also an important tool to prove universality in the bulk of eigenvalues for Wigner random matrices with matrix entries following so-called *Gaussian divisible distributions*. The generalisation from these special distributions to the general case, however, uses Lindeberg's idea of telescoping sums; see Johansson [17] and Erdős et al. [13].

Now, assume that X and Z are independent and define the interpolating sequence $Y_t = \sqrt{t}X + \sqrt{1-t}Z$, $0 \le t \le 1$. Note that, if $\mathbb{E}X = \mathbb{E}Z = 0$, then $\mathbb{E}Y_t = 0$ and, if $\operatorname{Cov}(X) = \operatorname{Cov}(Z)$, then $\operatorname{Cov}(Y_t) = \operatorname{Cov}(X)$ for all t (which may serve as an explanation why this particular Y_t is a good choice). With h_i being the partial derivative in the ith coordinate, we can write

$$\begin{aligned} \mathbb{E}h(X) - \mathbb{E}h(Z) &= \int_0^1 \frac{\partial}{\partial t}\mathbb{E}h(Y_t)\,\mathrm{d}t \\ &= \frac{1}{2}\int_0^1 \mathbb{E}\biggl\{\frac{1}{\sqrt{t}}\sum_i X_i h_i(Y_t) - \frac{1}{\sqrt{1-t}}\sum_i Z_i h_i(Y_t)\biggr\}\,\mathrm{d}t \end{aligned} \tag{2.3}$$

(differentiation in (2.3) corresponds to taking differences in (2.2) and integration replaces summation, but this is only a technical difference). One can easily see that, on the right-hand side of (2.3), the coordinates are treated symmetrically. The result obtained by Slepian [25] (known as *Slepian's Lemma*) is valid under the assumption that X and Z are centred Gaussian vectors having a different covariance structure. In this case, the Gaussian integration by parts formula

$$\mathbb{E}\bigl\{Z_i h_i(Z)\bigr\} = \sum_{j=1}^n \operatorname{Cov}(Z_i, Z_j)\mathbb{E}h_{ij}(Z) \tag{2.4}$$

can be used to estimate the error on the right-hand side of (2.3) in terms of the covariances. Stein [26], on the other hand, considered the univariate case, but where X is not Gaussian. Although (2.4) can still be used for Z, it needs to be replaced by an approximate version of (2.4) for X.

In order to formalise this approximate version of the Gaussian integration by parts formula, we will make use of a multivariate generalisation of *Stein couplings*, which were introduced by Chen and Röllin [10] in the univariate case, and then give more concrete constructions later on. Throughout this article summations will always range from 1 to n unless otherwise stated.

Definition 2.1. *Let (X, X', G) be a triple of n-dimensional random vectors. We say that the triple is a Stein coupling if, for any smooth enough function $f:\mathbb{R}^n \to \mathbb{R}$, we have*

$$\mathbb{E}\sum_i X_i f_i(X) = \mathbb{E}\sum_i G_i\bigl(f_i\bigl(X'\bigr) - f_i(X)\bigr) \tag{2.5}$$

whenever the involved expectations exist.

Remark 2.2. *If* (X, X', G) *is a Stein coupling, it follows from the definition that*

$$\mathbb{E}X_i = 0, \qquad \mathbb{E}(G_i D_j + G_j D_i) = 2\operatorname{Cov}(X_i, X_j) \tag{2.6}$$

for all i *and* j, *where we let* $D = X' - X$ *throughout this article (apply* (2.5) *to the functions* $f(x) = x_i$ *and* $f(x) = x_i x_j$, *respectively). If* (2.5) *is replaced by the stronger condition that*

$$\mathbb{E}\{X_i f_i(X)\} = \mathbb{E}\{G_i(f_i(X') - f_i(X))\} \tag{2.7}$$

for all i, *then*

$$\mathbb{E}(G_i D_j) = \mathbb{E}(G_j D_i) = \operatorname{Cov}(X_i, X_j) \tag{2.8}$$

for all i *and* j.

Equation (2.5) is the key condition to obtain an approximate Gaussian integration by parts formula: if X and X' are close to each other, then the difference on the right-hand side of (2.5) can be approximated by the corresponding derivatives, leading to a formula similar to (2.4). Hence, it is crucial that X' is only a *small perturbation* of X.

The following result, although not difficult to prove, is crucial for our approach. On one hand, it measures how closely X satisfies the Gaussian integration by parts formula and, on the other hand, also compares the covariances of X and Z (which in this article we will mostly assume to be the same). To make things more transparent, we keep everything explicit in terms of the function h, instead of using the usual approach via Stein equation and its solution.

Unless otherwise stated, we will assume throughout this article that

$$\mathbb{E}X = 0, \qquad \operatorname{Var} X_i = \sigma_i^2, \qquad \mathbb{E}|X_i|^3 = \tau_i^3 < \infty, \qquad \bar{\tau} = \sup_i \tau_i. \tag{2.9}$$

We will denote by $\Sigma = (\sigma_{ij})_{1\le i,j\le n}$ the covariance matrix of X, where $\sigma_{ij} = \mathbb{E}(X_i X_j)$, and we have $\sigma_{ii} = \sigma_i^2$.

Lemma 2.1. *Let* (X, X', G) *be a Stein coupling. Let* X'' *and* $\tilde{D}$ *be* n*-dimensional random vectors and let* S *be a random* $n \times n$ *matrix. Define* $D = X' - X$ *and* $D' = X'' - X$. *Assume that, for all* k *and* l,

$$\mathbb{E}(G_k D_l | X) = \mathbb{E}(G_k \tilde{D}_l | X), \qquad \mathbb{E}(S_{kl} | X) = \sigma_{kl}. \tag{2.10}$$

Let $Z \sim \mathrm{MVN}_n(0, \Sigma)$ *be independent of the previous random vectors. Then, for any three times partially differentiable function* h,

$$\begin{aligned}\mathbb{E}h(X) - \mathbb{E}h(Z) = {}& \frac{1}{2}\int_0^1 \mathbb{E}R_1(t)\,\mathrm{d}t - \frac{1}{2}\int_0^1\int_0^1 t^{1/2}\mathbb{E}R_2(t,s)\,\mathrm{d}s\,\mathrm{d}t \\ & + \frac{1}{2}\int_0^1\int_0^1\int_0^1 s t^{1/2}\mathbb{E}R_3(t, sr)\,\mathrm{d}r\,\mathrm{d}s\,\mathrm{d}t,\end{aligned} \tag{2.11}$$

where

$$R_1(t) = \sum_{k,l}(G_k\tilde{D}_l - S_{kl})h_{kl}\big(\sqrt{t}X'' + \sqrt{1-t}Z\big),$$

$$R_2(t,u) = \sum_{k,l,m}(G_k\tilde{D}_l - S_{kl})D'_m h_{klm}\big(\sqrt{t}X + u\sqrt{t}D' + \sqrt{1-t}Z\big),$$

$$R_3(t,u) = \sum_{k,l,m} G_k D_l D_m h_{klm}\big(\sqrt{t}X + u\sqrt{t}D + \sqrt{1-t}Z\big),$$

provided that $\mathbb{E}R_i(\cdot)$ *exists for* $i = 1, 2, 3$. *In particular,*

$$\big|\mathbb{E}h(X) - \mathbb{E}h(Z)\big| \le \frac{1}{2}\sup_t\big|\mathbb{E}R_1(t)\big| + \frac{1}{3}\sup_{t,s}\big|\mathbb{E}R_2(t,s)\big| + \frac{1}{6}\sup_{t,s}\big|\mathbb{E}R_3(t,s)\big|.$$

It seems rather difficult at this point to convey the purpose of all the random vectors appearing in the lemma. Probably the best way to get an intuition for such couplings is to go through the different applications given later on; we also refer to Chen and Röllin [10] for the univariate case, where further examples are discussed. We note that finding the appropriate random vectors will usually require some trial and error.

Remark 2.3. *Let us make a few comments at this point.*

1. *We will use the following simple fact in the applications. If (X, X', G) is a Stein coupling satisfying the stronger condition* (2.7) *and if there is a σ-algebra $\mathcal{F}'' \supset \sigma(X'')$ such that*
$$\mathbb{E}\big(G_k \tilde{D}_l | \mathcal{F}''\big) = \mathbb{E}\big(S_{kl} | \mathcal{F}''\big), \tag{2.12}$$
then $\mathbb{E}R_1(t) = 0$.
2. *Except for the case of local dependence in Section* 3.5, *we will choose $\tilde{D} = D$.*
3. *The result can be easily extended to include other error terms from the proof of the lemma under weaker conditions. We will use the following extension later on. If (X, X', G) is not a Stein coupling, then one can include a measure of how close* (2.5) *is satisfied; with*
$$\begin{aligned} R_0(t) = \sum_k \big\{ & X_k h_k\big(\sqrt{t}X + \sqrt{1-t}Z\big) - G_k h_k\big(\sqrt{t}X' + \sqrt{1-t}Z\big) \\ & + G_k h_k\big(\sqrt{t}X + \sqrt{1-t}Z\big)\big\}, \end{aligned}$$
an additional $\frac{1}{2}\int_0^1 \frac{1}{\sqrt{t}}\mathbb{E}R_0(t)\,\mathrm{d}t$ appears on the right-hand side of (2.11).
4. *If (X, X', G) is not a Stein coupling, the identities* (2.6) *and* (2.8) *are no longer valid and need to be replaced by corresponding approximate versions.*
5. *Note that the difference $|G_k \tilde{D}_l - S_{kl}|$ in $R_2(t, u)$ can usually be estimated by $|G_k \tilde{D}_l| + |S_{kl}|$ without changing the rates of convergence. This is not the case for $R_1(t)$, where more care is required.*

Proof of Lemma 2.1. Define the interpolating sequence $Y_t = \sqrt{t}X + \sqrt{1-t}Z$, $0 \le t \le 1$. Starting from (2.3), and using (2.4) and (2.5), we obtain

$$\begin{aligned} \mathbb{E}h(X) - \mathbb{E}h(Z) &= \int_0^1 \frac{\partial}{\partial t}\mathbb{E}h(Y_t)\,\mathrm{d}t \\ &= \frac{1}{2}\int_0^1 \mathbb{E}\left\{\sum_k \frac{1}{\sqrt{t}} X_k h_k(Y_t) - \sum_k \frac{1}{\sqrt{1-t}} Z_k h_k(Y_t)\right\}\mathrm{d}t \\ &= \frac{1}{2}\int_0^1 \mathbb{E}\left\{\sum_k \frac{1}{\sqrt{t}} G_k\big(h_k\big(Y_t'\big) - h_k(Y_t)\big) - \sum_{k,l} \sigma_{kl} h_{kl}(Y_t)\right\}\mathrm{d}t, \end{aligned} \tag{2.13}$$

where $Y_t' = \sqrt{t}X' + \sqrt{1-t}Z$. Let us recall the definition of $R_1(t)$ and introduce two additional error terms:

$$R_1(t) = \sum_{k,l} (G_k \tilde{D}_l - S_{kl}) h_{kl}\big(Y_t''\big),$$

$$R_4(t) := \sum_{k,l} (S_{kl} - \sigma_{kl}) h_{kl}(Y_t), \qquad R_5(t) := \sum_{k,l} G_k (D_l - \tilde{D}_l) h_{kl}(Y_t),$$

where $Y_t'' = \sqrt{t}X'' + \sqrt{1-t}Z$. Applying

$$h_k\big(Y_t'\big) - h_k(Y_t) = \int_0^1 \sqrt{t} \sum_l D_l h_{kl}\big(Y_t + s\sqrt{t}D\big)\,\mathrm{d}s$$

to (2.13), and adding and subtracting the terms from $R_1(t)$, $R_4(t)$ and $R_5(t)$ yields

$$
\begin{aligned}
&\mathbb{E}h(X) - \mathbb{E}h(Z) \\
&\quad = \frac{1}{2}\int_0^1 \mathbb{E}\biggl\{\int_0^1 \sum_{k,l} G_k D_l h_{kl}\bigl(Y_t + s\sqrt{t}D\bigr)\,\mathrm{d}s - \sum_{k,l}\sigma_{kl} h_{kl}(Y_t)\biggr\}\,\mathrm{d}t \\
&\quad = \frac{1}{2}\int_0^1 \mathbb{E}\bigl\{R_1(t) + R_4(t) + R_5(t)\bigr\}\,\mathrm{d}t \\
&\qquad + \frac{1}{2}\int_0^1 \mathbb{E}\biggl\{\sum_{k,l}(G_k\tilde{D}_l - S_{kl})\bigl(h_{kl}(Y_t) - h_{kl}\bigl(Y_t''\bigr)\bigr)\biggr\}\,\mathrm{d}t \\
&\qquad + \frac{1}{2}\int_0^1\int_0^1 \mathbb{E}\biggl\{\sum_{k,l} G_k D_l\bigl(h_{kl}\bigl(Y_t + s\sqrt{t}D\bigr) - h_{kl}(Y_t)\bigr)\biggr\}\,\mathrm{d}s\,\mathrm{d}t.
\end{aligned}
$$

Note that, under (2.10), $\mathbb{E}R_4(t) = \mathbb{E}R_5(t) = 0$. Taylor expansion in the last two lines yields the final result; we refer to Chen and Röllin [10] for a more detailed exposition of the proof in the univariate case. □

We now derive general norm bounds from Lemma 2.1, along the lines of Raič [20], Chatterjee and Meckes [8] and Meckes [18]. Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and let $|||\cdot|||$ be a norm on $\mathbb{R}^{n\times n}$, the space of $n\times n$ matrices. Define the following measures of smoothness of h. For $k\geq 1$, let

$$
M_k(h) = \sup_{x\in\mathbb{R}^n}\sup_{u^{(1)},\dots,u^{(k)}\in\mathbb{R}^n}\sum_{i_1,\dots,i_k=1}^n \frac{u_{i_1}^{(1)}\cdots u_{i_k}^{(k)}}{\|u^{(1)}\|\cdots\|u^{(k)}\|} h_{i_1,\dots,i_k}(x),
$$

and for $k\geq 2$ define

$$
\widetilde{M}_k(h) = \sup_{x\in\mathbb{R}^n}\sup_{A\in\mathbb{R}^{n\times n}}\sup_{u^{(3)},\dots,u^{(k)}\in\mathbb{R}^n}\sum_{i_1,\dots,i_k=1}^n \frac{A_{i_1 i_2}u_{i_3}^{(3)}\cdots u_{i_k}^{(k)}}{|||A|||\,\|u^{(3)}\|\cdots\|u^{(k)}\|} h_{i_1,\dots,i_k}(x)
$$

(if $k=2$, the third supremum in the definition of $\widetilde{M}_k(h)$ is just ignored). We then have the following straitforward result.

Lemma 2.2. *Under the assumptions of Lemma 2.1, let $\mathcal{F}''$ be a σ-algebra with $\sigma(X'')\subset\mathcal{F}''$. Then, for all $0\leq t, s\leq 1$,*

$$
\begin{aligned}
\bigl|\mathbb{E}R_1(t)\bigr| &\leq \widetilde{M}_2(h)\mathbb{E}\bigl\{\bigl|\bigl|\bigl|\mathbb{E}\bigl(G\tilde{D}^t - S|\mathcal{F}''\bigr)\bigr|\bigr|\bigr|\bigr\}, \\
\bigl|\mathbb{E}R_2(t,s)\bigr| &\leq \widetilde{M}_3(h)\mathbb{E}\bigl\{\bigl(\bigl|\bigl|\bigl|G\tilde{D}^t\bigr|\bigr|\bigr| + |||S|||\bigr)\bigl\|D'\bigr\|\bigr\}, \\
\bigl|\mathbb{E}R_3(t,s)\bigr| &\leq M_3(h)\mathbb{E}\bigl\{\|G\|\|D\|^2\bigr\}.
\end{aligned}
$$

It is clear from this lemma that the optimal choice of the norms $\|\cdot\|$ and $|||\cdot|||$ very much depends on the involved random vectors and how they are coupled. This, in turn, determines which functions h are considered smooth enough to yield informative bounds.

Let us fix some notation before we proceed. We denote by $\|\cdot\|_\infty$ the supremum norm of functions. For $k\geq 1$ and a k-times partially differentiable function $f:\mathbb{R}^n\to\mathbb{R}$, we let

$$
|f|_k = \sup_{1\leq i_1\leq\cdots\leq i_k\leq n}\|f_{i_1\dots i_k}\|_\infty.
$$

For functions $g:\mathbb{R}\to\mathbb{R}$ we will use the notation $\|g'\|_\infty, \|g''\|_\infty, \dots,$ instead of the equivalent $|g|_1, |g|_2, \dots.$

The couplings we construct in this article are such that the random vectors and matrices in Lemma 2.2 are small with respect to the L_1-norms

$$\|u\|_1 = \sum_{i=1}^{n} |u_i| \quad \text{and} \quad |||A|||_1 = \sum_{i,j=1}^{n} |A_{ij}|.$$

It is not difficult to see that with respect to these norms we have

$$M_k(h) = \widetilde{M}_k(h) = |h|_k.$$

For this reason, we will directly formulate our results in terms of $|h|_k$.

Note that this is in contrast to the results for multivariate normal approximation of Chatterjee and Meckes [8] and Reinert and Röllin [21]. There, the vectors and matrices are typically closer in L_2 than in L_1. Meckes [18] showed that in this case $|\cdot|_k$ is too strong to measure the smoothness of h, resulting in suboptimal dependence on the dimension. Using instead $M_k(h)$ and $\widetilde{M}_k(h)$ with respect to the L_2-norms, Meckes [18] showed that the dependence on the dimension can be substantially reduced.

Remark 2.4. *One may be interested in comparing the distributions of $f(X)$ and $f(Z)$ for some specific function $f:\mathbb{R}^n \to \mathbb{R}$. To this end, choose $h(x) = g(f(x))$ for $g:\mathbb{R}\to\mathbb{R}$. Then, if (1.1) is small for all three times differentiable functions g, then we can conclude that $f(X)$ and $f(Z)$ are close in distribution. We record the useful estimates*

$$|h|_1 \le |f|_1 \|g'\|_\infty, \qquad |h|_2 \le |f|_2 \|g'\|_\infty + |f|_1^2 \|g''\|_\infty,$$
$$|h|_3 \le |f|_3 \|g'\|_\infty + 3|f|_1|f|_2\|g''\|_\infty + |f|_1^3 \|g'''\|_\infty.$$

Remark 2.5. *A particular function of interest is*

$$f(x) = \log \sum_{p=1}^{m} \mathrm{e}^{\beta y^{(p)}(x)}$$

for functions $y^{(p)}:\mathbb{R}^n \to \mathbb{R}$, $1 \le p \le m$. Define $\gamma_k = \sup_p |y^{(p)}|_k$; it is straightforward to check that

$$|f|_1 \le \beta\gamma_1, \qquad |f|_2 \le \beta\gamma_2 + 2\beta^2\gamma_1^2, \qquad |f|_3 \le \beta\gamma_3 + 6\beta^2\gamma_1\gamma_2 + 6\beta^3\gamma_1^3.$$

3. Couplings

Many of the Stein couplings discussed by Chen and Röllin [10] can be adapted to the multivariate case: exchangeable pairs, size-biasing, local dependence, etc. Instead of generalising all of them here (which will be done elsewhere with emphasis on multivariate normal approximation for fixed dimension) we only go through a few of them explicitly and instead present some other couplings not discussed by Chen and Röllin [10].

3.1. *A theoretical result*

One may wonder if, given a pair (X, X') with $\mathbb{E}X = 0$, there exists a G to make the triple (X, X', G) a Stein coupling. This question has been answered by Chen and Röllin [10] for the univariate case, but the construction given there can also be used in the multivariate setting. Let $\mathcal{F} = \sigma(X)$ be the σ-algebra induced by X and let $\mathcal{F}' = \sigma(X')$. Define formally the sequence

$$G = -X + \mathbb{E}(X|\mathcal{F}') - \mathbb{E}(\mathbb{E}(X|\mathcal{F}')|\mathcal{F}) + \mathbb{E}(\mathbb{E}(\mathbb{E}(X|\mathcal{F}')|\mathcal{F})|\mathcal{F}') - \cdots.$$

If the sequence converges absolutely in each coordinate, then this will make (X, X', G) a Stein coupling. Indeed, $\mathbb{E}(G|\mathcal{F}) = -X$ and $\mathbb{E}(G|\mathcal{F}') = 0$ so that (2.5) is satisfied.

To motivate the choice of G used in the next few settings, consider the case where the coordinates of X are independent. Let I be uniformly distributed on $\{1,\dots,n\}$, independent of all else. Define the vector $X^{(i)}$ by

$$X_k^{(i)} = (1-\delta_{ki})X_k,$$

where δ_{ij} is the Dirac delta function. Let $X' = X^{(I)}$; that is, X' is the vector where we have set a randomly chosen coordinate to 0. Denote by e_i the unit vector in direction i. Using independence of the coordinates,

$$\mathbb{E}\bigl(X|\mathcal{F}'\bigr) = \mathbb{E}\bigl(X^{(I)} + e_I X_I | X^{(I)}\bigr) = X^{(I)}$$

and

$$\mathbb{E}\bigl(X'|\mathcal{F}\bigr) = \frac{1}{n}\sum_i X^{(i)} = \biggl(1-\frac{1}{n}\biggr)X.$$

Hence,

$$\begin{aligned} G &= -X + X' - \biggl(1-\frac{1}{n}\biggr)X + \biggl(1-\frac{1}{n}\biggr)X' - \biggl(1-\frac{1}{n}\biggr)^2 X + \biggl(1-\frac{1}{n}\biggr)^2 X' + \cdots \\ &= -e_I X_I - \biggl(1-\frac{1}{n}\biggr)e_I X_I - \biggl(1-\frac{1}{n}\biggr)^2 e_I X_I - \cdots = -ne_I X_I. \end{aligned}$$

3.2. *Independent coordinates*

In order to illustrate the method in a simple setting, we start with independent coordinates using (X, X', G) derived in the previous section.

Theorem 3.1. *Let X be as in (2.9) and assume the coordinates of X are independent. If Z is a vector of independent centred Gaussian random variables with the same variances as X, then*

$$\bigl|\mathbb{E}h(X) - \mathbb{E}h(Z)\bigr| \le \frac{5}{6}\sum_i \tau_i^3 \|h_{iii}\|_\infty.$$

Proof. Let $G_i = -n\delta_{iI}X_i$ and $X' = X'' = X^{(I)}$, hence $D_i = D_i' = -\delta_{iI}X_i$. Let $\tilde{D} = D$ and $S_{ij} = n\sigma_i^2\delta_{iI}\delta_{jI}$. It is easy to see that (X, X', G) is a Stein coupling satisfying the stronger condition (2.7), that (2.10) is satisfied and that (2.12) holds with $\mathcal{F}'' = \sigma(X'', I)$; hence $\mathbb{E}R_1(t) = 0$. The following estimates are immediate:

$$\begin{aligned} \bigl|R_2(t)\bigr| &\le \sum_i \|h_{iii}\|_\infty\bigl(\sigma_i^2\mathbb{E}|X_i| + \mathbb{E}|X_i|^3\bigr) \le 2\sum_i \|h_{iii}\|_\infty \mathbb{E}|X_i|^3, \\ \bigl|R_3(t)\bigr| &\le \sum_i \|h_{iii}\|_\infty \mathbb{E}|X_i|^3. \end{aligned}$$

Lemma 2.1 concludes the theorem. □

Using Lindeberg's telescoping sum and Taylor expansion, and noting that the first two moments of X and Z match, one easily obtains

$$\begin{aligned} &\bigl|\mathbb{E}h(X) - \mathbb{E}h(Z)\bigr| \\ &\quad \le \frac{1}{6}\sum_i \bigl(\mathbb{E}|X_i|^3 + \mathbb{E}|Z_i|^3\bigr)\|h_{iii}\|_\infty \le \frac{(1+\sqrt{8/\pi})}{6}\sum_i \tau_i^3 \|h_{iii}\|_\infty. \end{aligned}$$

Not surprisingly, the constants obtained via Stein's method are larger for the case of independent random variables. However, applications with dependencies is the main purpose of using Stein's method.

3.3. *Weak dependence*

A simple way to measure how much a single coordinate X_i is influenced by the other coordinates is to look at the fluctuation of the conditional mean and variance of X_i. To this end, let X be as in (2.9) and define $X^{(i)}$ as in Section 3.1. Furthermore, let

$$\mu_i\big(X^{(i)}\big) = \mathbb{E}\big(X_i | X^{(i)}\big), \qquad \sigma_i^2\big(X^{(i)}\big) = \operatorname{Var}\big(X_i | X^{(i)}\big).$$

Then we have the following.

Theorem 3.2. *Let X be as in (2.9) and let $Z \sim \mathrm{MVN}_n(0, \Sigma)$. Then*

$$\big|\mathbb{E}h(X) - \mathbb{E}h(Z)\big| \le \sum_i \|h_i\|_\infty \mathbb{E}\big|\mu_i\big(X^{(i)}\big)\big| + \frac{1}{2}\sum_i \|h_{ii}\|_\infty \big(\mathbb{E}\mu_i\big(X^{(i)}\big)^2 + \mathbb{E}\big|\sigma_i^2\big(X^{(i)}\big) - \sigma_i^2\big|\big) + \frac{5}{6}\sum_i \tau_i^3 \|h_{iii}\|_\infty.$$

Proof. Define G, X', X'', S as in the proof of Theorem 3.1; the error terms R_2 and R_3 can be bounded in the same way. As (X, X', G) is not necessarily a Stein coupling, we need the additional error term

$$\mathbb{E}R_0(t) = \mathbb{E}\sum_i X_i h_i\big(\sqrt{t}X^{(i)} + \sqrt{1-t}Z\big) = \mathbb{E}\sum_i \mu_i\big(X^{(i)}\big) h_i\big(\sqrt{t}X^{(i)} + \sqrt{1-t}Z\big)$$

(see Remark 2.3). Furthermore,

$$\begin{aligned}\mathbb{E}R_1(t) &= \mathbb{E}\sum_i \big(X_i^2 - \sigma_i^2\big) h_{ii}\big(\sqrt{t}X^{(i)} + \sqrt{1-t}Z\big) \\ &= \mathbb{E}\sum_i \big(\big(X_i - \mu_i\big(X^{(i)}\big)\big)^2 - \sigma_i^2 - \mu_i\big(X^{(i)}\big)^2\big) h_{ii}\big(\sqrt{t}X^{(i)} + \sqrt{1-t}Z\big).\end{aligned}$$

This easily leads to the final bound. □

Note that if the X_i are independent, Theorem 3.2 reduces to Theorem 3.1. Götze and Tikhomirov [14] assumed that $\mu_i(X^i) = 0$ almost surely to obtain convergence rates to the semi-circular law in random matrix theory under such dependence.

3.4. *Constant sum and symmetry*

Recall the classic occupancy problem from the Introduction. The sum of the vector that describes the number of balls in each urn is equal to the total number of balls and hence, itself, does not satisfy a central limit theorem. This motivates us to consider general centered vectors X that satisfy

$$\sum_i^n X_i = 0 \tag{3.1}$$

almost surely.

To apply our method, we will need to make more assumptions. A random vector $X = (X_1, \dots, X_n)$ is called exchangeable if its distribution is invariant under permutation of the coordinates. Note that (3.1) implies $\sum_j \sigma_{ij} = 0$ for each i, and combined with exchangeability, we therefore have

$$\sigma_{ij} = -\frac{\sigma_1^2}{n-1} \tag{3.2}$$

for all $i \ne j$.

Theorem 3.3. *Let X be an exchangeable random vector satisfying* (2.9) *and let $Z \sim \mathrm{MVN}_n(0, \Sigma)$. Then*

$$\bigl|\mathbb{E}h(X) - \mathbb{E}h(Z)\bigr| \le |h|_2\biggl[\mathrm{Var}\biggl(\sum_i X_i^2\biggr)\biggr]^{1/2} + 16|h|_3 n\tau_1^3. \tag{3.3}$$

Remark 3.1. *Note that the theorem can also be applied if X is not exchangeable, but h symmetric instead, that is if $h(x)$ remains the same under any permutation of the coordinates. In that case, Theorem* 3.3 *can be applied to the randomly permuted X. Note that τ_1^3 is then replaced by $n^{-1}\sum_i \tau_i^3$ for the final result.*

Proof of Theorem 3.3. For $x \in \mathbb{R}^n$, let $x^{ik} \in \mathbb{R}^n$ be the vector obtained by interchanging the ith and kth coordinate of x (if $i = k$ then $x^{ik} = x$). Note that due to exchangeability,

$$\mathbb{E}\bigl\{\varphi(X_i)h_i\bigl(X^{ik}\bigr)\bigr\} = \mathbb{E}\bigl\{\varphi(X_k)h_i(X)\bigr\} \tag{3.4}$$

for any function φ for which the expectations exist. Furthermore, for $(i, j, k, l) \in [n]^4$ with

$$i = k \quad \Longleftrightarrow \quad j = l, \tag{3.5}$$

denote by x^{ijkl} a permutation of x such that

$$\mathbb{E}\bigl\{\varphi(X_j, X_l)h_{ik}(X)\bigr\} = \mathbb{E}\bigl\{\varphi(X_i, X_k)h_{ik}\bigl(X^{ijkl}\bigr)\bigr\} \tag{3.6}$$

for all functions φ for which the expectations exist. Note that this permutation can be defined independently of x and h: if X is exchangeable, keep $[n] \setminus \{i, j, k, l\}$ fixed, map $j \mapsto i$ and $l \mapsto k$ and map the remaining numbers among each other in any arbitrary, but fixed way. Let (I, J, K, L) be distributed on $[n]^4$, such that (I, J, K, L) is uniform on $[n]^3$ and, given (I, J, K), L is uniform on $[n] \setminus \{J\}$ if $I \neq K$, and $J = L$ if $I = K$; hence, (I, J, K, L) satisfies (3.5). Define

$$X' := X^{IK}, \qquad X'' := X^{IJKL}$$

and

$$G_k = -n\delta_{kI}X_k, \qquad \tilde{D}_k = D_k, \qquad S_{kl} = n^2\delta_{kI}\delta_{lK}\sigma_{kl};$$

note that

$$D_l = \delta_{lI}(X_K - X_I) + \delta_{lK}(X_I - X_K), \qquad D'_l = \sum_{m\in\{I,J,K,L\}} \delta_{lm}\bigl(X''_l - X_l\bigr).$$

Fix t and let, for notational convenience, $f_\cdot(x) = \mathbb{E}h_\cdot(\sqrt{t}x + \sqrt{1-t}Z)$, where $\cdot$ stands for i, ij or ijk. Clearly, $\mathbb{E}\{G_k f_k(X)\} = \mathbb{E}\{X_k f_k(X)\}$. Using exchangeability of X, we can use (3.4) to obtain

$$\begin{aligned}\mathbb{E}\sum_k G_k f_k\bigl(X'\bigr) &= -n\mathbb{E}\bigl\{X_I f_I\bigl(X^{IK}\bigr)\bigr\} = -n\mathbb{E}\bigl\{X_K f_I(X)\bigr\}\\ &= -\frac{1}{n}\mathbb{E}\sum_{i,k} X_k f_i(X) = 0.\end{aligned}$$

Hence, (X, X', G) is a Stein coupling. Now,

$$\begin{aligned}\mathbb{E}R_1(t) &= \mathbb{E}\sum_{k,l}(S_{kl} - G_k D_l) f_{kl}\bigl(X''\bigr)\\ &= \mathbb{E}\sum_{k,l}\bigl[n^2\delta_{kI}\delta_{lK}\sigma_{kl} + \delta_{kI}nX_k\bigl(\delta_{lI}(X_K - X_l) + \delta_{lK}(X_I - X_l)\bigr)\bigr] f_{kl}\bigl(X''\bigr)\\ &= n^2\mathbb{E}\sigma_{IK} f_{IK}\bigl(X''\bigr) + n\mathbb{E}X_I(X_K - X_I) f_{II}\bigl(X''\bigr) + n\mathbb{E}X_I(X_I - X_K) f_{IK}\bigl(X''\bigr).\end{aligned}$$

Using exchangeability and (3.2),

$$\begin{aligned} n^2\mathbb{E}\{\sigma_{IK} f_{IK}(X'')\} &= n^2\mathbb{E}\{\sigma_{JL} f_{IK}(X)\} \\ &= \frac{1}{n}\mathbb{E}\sum_{i,j}\sigma_{jj} f_{ii}(X) + \frac{1}{n(n-1)}\mathbb{E}\sum_{i,j,k\neq i,l\neq j}\sigma_{jl} f_{ik}(X) \\ &= \sigma_1^2\mathbb{E}\sum_i f_{ii}(X) - \frac{\sigma_1^2}{n-1}\mathbb{E}\sum_{i,k\neq i} f_{ik}(X). \end{aligned}$$

Furthermore, using (3.1) and (3.6),

$$\begin{aligned} n\mathbb{E}\{X_I(X_K - X_I) f_{II}(X'')\} &= n\mathbb{E}\{X_J(X_L - X_J) f_{II}(X)\} \\ &= \frac{1}{n^2}\mathbb{E}\sum_{i,j,l} X_j(X_l - X_j) f_{ii}(X) \\ &= -\frac{1}{n}\mathbb{E}\sum_j X_j^2\sum_i f_{ii}(X) \end{aligned}$$

and

$$\begin{aligned} & n\mathbb{E}\{X_I(X_I - X_K) f_{IK}(X'')\} \\ &\quad = n\mathbb{E}\{X_J(X_J - X_L) f_{IK}(X)\} \\ &\quad = \frac{1}{n^2(n-1)}\mathbb{E}\sum_{i,j,k\neq i,l\neq j} X_j(X_j - X_l) f_{ik}(X) \\ &\quad = \frac{1}{n^2}\mathbb{E}\sum_{i,j,k\neq i} X_j^2 f_{ik}(X) - \frac{1}{n^2(n-1)}\mathbb{E}\sum_{i,j,k\neq i,l\neq j} X_j X_l f_{ik}(X) \\ &\quad = \frac{1}{n^2}\mathbb{E}\sum_j X_j^2\sum_{i,k\neq i} f_{ik}(X) - \frac{1}{n^2(n-1)}\mathbb{E}\sum_{j,l\neq j} X_j X_l\sum_{i,k\neq i} f_{ik}(X) \\ &\quad = \frac{1}{n^2}\mathbb{E}\sum_j X_j^2\sum_{i,k\neq i} f_{ik}(X) + \frac{1}{n^2(n-1)}\mathbb{E}\sum_j X_j^2\sum_{i,k\neq i} f_{ik}(X) \\ &\quad = \frac{1}{n(n-1)}\mathbb{E}\sum_j X_j^2\sum_{i,k\neq i} f_{ik}(X), \end{aligned}$$

where for the last equality we used that $\frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}$. Hence,

$$\begin{aligned} |\mathbb{E}R_1(t)| &\le \frac{1}{n(n-1)}\left|\mathbb{E}\left\{\left(\sum_j X_j^2 - n\sigma_1^2\right)\sum_{i,k\neq i} f_{ik}(X)\right\}\right| \\ &\quad + \frac{1}{n}\left|\mathbb{E}\left\{\left(\sum_j X_j^2 - n\sigma_1^2\right)\sum_i f_i(X)\right\}\right| \\ &\le 2|h|_2\left[\operatorname{Var}\left(\sum_i X_i^2\right)\right]^{1/2}. \end{aligned}$$

This gives the first part of the result. Now,

$$\begin{aligned}\mathbb{E}R_2(t,u) &= \mathbb{E}\sum_{k,l,m}(S_{kl} - G_k\tilde{D}_l)D'_m f_{klm}\bigl(X+uD'\bigr)\\ &= n^2\mathbb{E}\sum_{m\in\{I,J,K,L\}}\sigma_{IK}\bigl(X''_m - X_m\bigr)f_{IKm}\bigl(X+uD'\bigr)\\ &\quad - n\mathbb{E}\sum_{l\in\{I,K\},m\in\{I,J,K,L\}}X_I\bigl(X'_l - X_l\bigr)\bigl(X''_m - X_m\bigr)f_{Ilm}\bigl(X+uD'\bigr)\end{aligned}$$

hence

$$\bigl|\mathbb{E}R_2(t,u)\bigr| \le 8|h|_3\tau_1\sum_{i,j}|\sigma_{ij}| + 32|h|_3 n\tau_1^3 \le 8|h|_3\tau_1 n\sigma_1^2 + 32|h|_3 n\tau_1^3.$$

Similarly,

$$\begin{aligned}\mathbb{E}R_3(t,u) &= \sum_{k,l,m}G_k D_l D_m h_{klm}(X+uD)\\ &= n\mathbb{E}\sum_{l\in\{I,K\},m\in\{I,K\}}X_I\bigl(X'_l - X_l\bigr)\bigl(X'_m - X_m\bigr)h_{klm}(X+uD),\end{aligned}$$

hence

$$\bigl|\mathbb{E}R_3(t,u)\bigr| \le 16|h|_3 n\tau_1^3. \qquad \square$$

3.5. *Local dependence*

Stein couplings to handle local dependence has already been discussed by Chen and Röllin [10], based on similar decompositions that appeared in many other places; we refer to the more detailed discussion in Chen and Röllin [10]. In particular, multivariate normal approximation for sums of locally dependent random vectors was considered by Rinott and Rotar' [23] and Raič [20].

Let $X = (X_1,\dots,X_n)$ be as in (2.9). Assume that, for each $i \in [n] := \{1,\dots,n\}$, there is a subset $A_i \subset [n]$ such that $X_{A_i^c}$ and X_i are independent. Assume further that for each $i \in [n]$ and $j \subset A_i$ there is a subset $B_{ij} \subset [n]$ such that $A_i \subset B_{ij}$ and $X_{B_{ij}^c}$ is independent of (X_i, X_j). Central limit theorems for sums of random variables satisfying this refined version of local dependence were analyzed in detail by Barbour et al. [2].

Theorem 3.4. *Let X as above. Let $Z \sim \mathrm{MVN}_n(0,\Sigma)$. Then, for any three times partially differentiable function h,*

$$\begin{aligned}\bigl|\mathbb{E}h(X) - \mathbb{E}h(Z)\bigr| &\le \frac{1}{3}\sum_i\sum_{j\in A_i}\sum_{k\in B_{ij}}\bigl(|\sigma_{ij}|\mathbb{E}|X_k| + \mathbb{E}|X_iX_jX_k|\bigr)\|h_{ijk}\|_\infty\\ &\quad + \frac{1}{6}\sum_i\sum_{j,k\in A_i}\mathbb{E}|X_iX_jX_k|\|h_{ijk}\|_\infty \le \frac{5}{6}\bar{\tau}^3 n\eta|h|_3,\end{aligned}$$

where $\eta = \sup_i\sum_{j\in A_i}|B_{ij}|$.

Proof. Let I be uniform on $[n]$ and, given I, let J be uniform on A_I. Define the vectors X', X'', G and $\tilde{D}$ and the matrix S as

$$\begin{aligned}&G_k = -\delta_{kI}nX_k, \qquad X'_k = \mathrm{I}[k\notin A_I]X_k, \qquad X''_k = \mathrm{I}[k\notin B_{IJ}]X_k,\\ &S_{kl} = n|A_I|\delta_{kI}\delta_{lJ}\sigma_{kl}, \qquad \tilde{D}_k = -|A_J|\delta_{kJ}X_k.\end{aligned}$$

Note that X' is independent of G, which makes (X, X', G) a Stein coupling satisfying the stronger condition (2.7), similarly as for the independent case. Furthermore, with $\mathcal{F}'' = \sigma(X'', I)$, (2.12) holds and therefore $\mathbb{E}R_1(t) = 0$. The final bound follows now easily from Lemma 2.1. □

As we can see from the case of the CLT, where $h(x) = g(\sum_i x_i)$, the typical scaling of X is such that $\bar{\tau}^3 \asymp n^{-3/2}$. With this scaling, a "typical" function h will have the property that $\mathbb{E}|h(X)| \asymp 1$, whereas the bound of Theorem 3.4 is of order $O(n^{-1/2})$.

Note that an m-dependent sequence is a special case of local dependence: we have $|A_i| = 1+2m$ and $B_{ij} \leq 1+3m$. However, the crucial aspect here is that the exact structure of the dependence is only important in terms of the size of A_i and B_{ij}. Any graph with maximal degree m that describes the dependence structure of X (that is, two subsets of vertices are independent if there is no edge between them) will have the upper bounds $|A_i| \leq 1+m$ and $|B_{ij}| \leq 1+2m$. In that case,

$$\eta \leq 2(m+1)^2. \tag{3.7}$$

4. Applications

In this section, we present two different types of applications. First, we consider concrete functions h, for which we determine under what kind of dependencies (1.1) is small. If we can control the first three derivatives of h, then we can analyse the universality of the given h with respect to dependence, for example for the different settings of the previous section. The first two applications below are of this type. We analyse universality with respect to local dependence only, but it is clear that many of the other settings can be used instead. In the case of local dependence, we are interested in how big the "neighbourhoods" A_i and B_{ij} are allowed to become while keeping the bounds on (1.1) small enough. We use η from Theorem 3.4 as a simple measure of neighbourhood size, and hence dependence. These applications are closely related to Chatterjee [6]. Whereas in the first application of the SK-model the dependence enters in a straightforward way, in the second application of last passage percolation on thin rectangles, an certain optimisation step has to be recalculated, including the measure of dependence η.

As a second type of application, we can consider more concrete vectors X, for which we want to show that (1.1) is small for a large class of functions h. In this situation, the structure of the dependence of X will either fit into one of the abstract settings of the previous section (this is the case for classic occupancy), or else, one has to construct a Stein coupling from scratch; the latter is the case for the Curie–Weiss model.

4.1. *Environment with dependencies in the Sherrington–Kirkpatrick spin glass model*

Consider the N-spin system $\{-1,1\}^N$. To each configuration $\sigma \in \{-1,1\}^N$ we assign the (random) Hamiltonian

$$H_N(\sigma) = \frac{\beta}{\sqrt{N}} \sum_{i<j} \xi_{ij}\sigma_i\sigma_j,$$

where $\xi = (\xi_{ij})_{1\leq i<j\leq n}$ is a family of random variables, which we call the *environment*. Given the environment ξ, we assign to each σ the probability

$$\mathbb{P}_N^{\xi}(\sigma) = \frac{e^{H_N(\sigma)}}{Z_N(\beta,\xi)},$$

where

$$Z_N(\beta,\xi) = \sum_{\sigma} e^{\beta H_N(\sigma)}.$$

Let

$$p_N(\beta) = \frac{1}{N}\mathbb{E}\log Z_N(\beta,\xi).$$

It was proved by Talagrand [28] that $p_N(\beta) \to p_\infty(\beta)$, the solution of the *Parisi formula*, if the ξ_{ij} are independent standard Gaussians. Carmona and Hu [5] showed that the same limit holds if the Gaussians are replaced by independent copies of any random variable ξ with $\mathbb{E}\xi = 0$ and $\mathbb{E}|\xi|^3 < \infty$. We shall extend this results to dependent environments. To this end define

$$\tilde{Z}_n(\beta, \xi) = \mathbb{E}^\xi\left\{e^{\beta \sum_{i=1}^n Y_i \xi_i}\right\},$$

where $Y_1, \dots, Y_n$ is any family of random variables such that Y_i only takes finitely many values and $|Y_i| \le 1$ for all i.

Lemma 4.1. *Let $\xi = (\xi_1, \dots, \xi_n)$ be a random environment such that $\mathbb{E}\xi_i = 0$, $\mathbb{E}\xi_i^2 = 1$ and $\mathbb{E}|\xi_i|^3 \le \bar{\tau}^3 < \infty$, satisfying the dependence structure of Theorem 3.4. Let $g \sim \mathrm{MVN}_n(0, \Sigma)$ where Σ is the covariance matrix of ξ. Then*

$$\left|\mathbb{E}\log \tilde{Z}_n(\beta, \xi) - \mathbb{E}\log \tilde{Z}_n(\beta, g)\right| \le 5\beta^3\bar{\tau}^3 n\eta. \tag{4.1}$$

Proof. Let $h(\xi) = \log \tilde{Z}_n(\beta, \xi)$; it is easy to see from Remark 2.5 that

$$\|h_{ijk}\| \le 6\beta^3$$

(note that $\gamma_2 = \gamma_3 = 0$ and $\gamma_1 \le 1$ as $|Y_i| \le 1$). Using Theorem 3.4, (4.1) is immediate. □

The following statement is a direct consequence of Lemma 4.1 for $n = N(N-1)/2$ and β replaced by $\beta N^{-1/2}$.

Theorem 4.2. *Assume the environment ξ satisfies the dependence structure of Theorem 3.4 with $\eta = \mathrm{o}(N^{1/2})$ and $\sigma_{ij} = 0$ for $i \ne j$. Then*

$$\frac{1}{N}\mathbb{E}\log Z_N(\beta, \xi) \to p_\infty(\beta).$$

Consider a fixed m-regular graph G on the set of vertices $V_N = \{(i, j)\colon 1 \le i < j \le N\}$. Let h_{ij} be i.i.d. centred random variables with finite third moments. Let

$$\xi_{ij} = \prod_{(k,l)\sim(i,j)} h_{kl}.$$

Then it is straightforward to see that these ξ_{ij} are centred and uncorrelated (note that ξ_{ij} does not contain h_{ij}). Clearly, from (3.7), $\eta \le 2(m+1)^2$ and hence we can apply Theorem 4.2 as long as $m = \mathrm{o}(N^{1/4})$. Noticing that (4.1) is independent of the underlying graph, we obtain the following.

Corollary 4.3. *Let G_N be a sequence of random m_N-regular graphs on V_N, where $m_N = \mathrm{o}(N^{1/4})$. Then, with ξ as above,*

$$\frac{1}{N}\mathbb{E}^{G_N}\log Z_N(\beta, \xi) \to p_\infty(\beta)$$

almost surely.

4.2. *Last passage percolation for thin rectangles*

The following statements about smooth approximation of the maximum function is well-known (and easy to verify).

Lemma 4.4. *Let m be a positive integer. For each $y \in \mathbb{R}^m$, let $f_0(y) = \max\{y_1, \dots, y_m\}$ and $f_\varepsilon(y) = \varepsilon \log \sum_i e^{y_i/\varepsilon}$. Then*

$$0 \le f_\varepsilon(y) - f_0(y) \le \varepsilon \log m.$$

Consider functions $y^{(p)}:\mathbb{R}^n \to \mathbb{R}$, $p = 1, \ldots, m$, and let

$$P_x = \max_{1\le p\le m} y^{(p)}(x). \tag{4.2}$$

The following theorem is similar to a result obtained by Chatterjee [6], but now includes η. To keep the bounds simple we make the stronger assumption that the functions $y^{(p)}$ are linear, which what we will need subsequently.

Theorem 4.5. *Let P_x be as above with linear functions $y^{(p)}$. Let X be a family of n centred random variables with finite third moments satisfying the dependence structure as in Theorem 3.4. Let $g:\mathbb{R}\to\mathbb{R}$ be three times differentiable. Then, for $Z \sim \mathrm{MVN}_n(0, \Sigma)$,*

$$\bigl|\mathbb{E}g(P_X) - \mathbb{E}g(P_Z)\bigr| \le \bigl(6\|g'\|_\infty + 6\|g''\|_\infty + \|g'''\|_\infty\bigr)n^{1/3}\eta^{1/3}\bar{\tau}\gamma_1 \log(m)^{2/3},$$

where $\gamma_1 = \sup_{1\le p\le m} |y^{(p)}|_1$.

Proof. Using the notation of Lemma 4.4, define the functions

$$h_0(x) = g\bigl(f_0(x)\bigr), \qquad h_\varepsilon(x) = g\bigl(f_\varepsilon(x)\bigr).$$

Clearly

$$\bigl|h_0(x) - h_\varepsilon(x)\bigr| \le \|g'\|_\infty \varepsilon \log m.$$

We now use Remark 2.5. We clearly have $\gamma_2 = \gamma_3 = 0$. Furthermore, using again Lemma 4.4, it is easy to check that,

$$\begin{aligned}|h_\varepsilon|_3 &\le |f_\varepsilon|_3\|g'\|_\infty + 3|f_\varepsilon|_1|f_\varepsilon|_2\|g''\|_\infty + |f_\varepsilon|_1^3\|g'''\|_\infty \\ &\le \varepsilon^{-2}\gamma_1^3\bigl(6\|g'\|_\infty + 6\|g''\|_\infty + \|g'''\|_\infty\bigr).\end{aligned}$$

Thus, using Theorem 3.4,

$$\begin{aligned}&\bigl|\mathbb{E}h_0(X) - \mathbb{E}h_0(Z)\bigr| \\ &\quad\le \|g'\|_\infty \varepsilon \log m + \bigl|\mathbb{E}h_\varepsilon(X) - \mathbb{E}h_\varepsilon(Z)\bigr| \\ &\quad\le \|g'\|_\infty \varepsilon \log m + C\varepsilon^{-2}\bar{\tau}^3 n\eta\gamma_1^3\bigl(6\|g'\|_\infty + 6\|g''\|_\infty + \|g'''\|_\infty\bigr).\end{aligned}$$

Choosing $\varepsilon = n^{1/3}\eta^{1/3}\log(m)^{-1/3}\bar{\tau}\gamma_1$, we obtain the final bound. □

Let us apply this result to last passage percolation on thin rectangles along the lines of Suidan [27]. Denote by π an increasing path from (1, 1) to (N, k) on the usual two dimensional lattice, where without loss of generality $k \le N$. Let

$$y^{(\pi)}(x) = \frac{k^{1/6}}{N^{1/2}}\biggl(\sum_{i\in\pi} x_i - 2\sqrt{Nk}\biggr)$$

and let P_x be as in (4.2), where the maximum ranges over all increasing paths π. Hence, P_x is the (standardized) longest increasing path between (1, 1) and (N, k), where each lattice point (i, j) contributes x_{ij} to the length of the path. If X is an i.i.d. family of geometric or exponential random variables, then Johansson [16] showed that the properly centred and standardized P_X will converge to F_2 (the Tracy–Widom distribution for Gaussian unitary ensembles) if $k = N$. For independent X_i that are neither exponentially nor geometrically distributed, the same results is only known for thin rectangles, that is, for k being of smaller order than N; see Bodineau and Martin [3], Baik and Suidan [1] and Suidan [27]. In particular, if X_i have finite third moments, then $k = \mathrm{O}(N^\alpha)$ for $\alpha < 1/7$. We shall expand this result to locally dependent X. If η remains bounded, we recover the same maximal order for k as in the independent case. If η grows with N, however, the maximal order of k be will be affected.

Corollary 4.6. *Let $X=(X_{ij})_{1\le i\le N,1\le j\le k}$ be a collection of $n=Nk$ random variables with mean* 0 *and variance* 1, *satisfying the dependence structure of Theorem* 3.4, *and let $Z\sim \mathrm{MVN}_n(0,\Sigma)$. Then, for any three times differentiable function $g:\mathbb{R}\to\mathbb{R}$*,

$$\left|\mathbb{E}g(P_X)-\mathbb{E}g(P_Z)\right|\le \frac{C(g,\bar\tau)\eta^{1/3}k^{7/6}\log(N)^{2/3}}{N^{1/6}}.$$

For some constant $C(g,\bar\tau)$. If $\sigma_{ij}=0$ for all $i\neq j$, then the P_X will converge to F_2 if $\bar\tau$ remains bounded and if

$$k=\mathrm{o}\bigl(N^{1/7}\log(N)^{-4/7}\eta^{-2/7}\bigr).$$

Proof. Clearly, $\gamma_1=k^{1/6}N^{-1/2}$. Furthermore,

$$m=\binom{N+k}{N}\le\left(\frac{N}{k}\right)^k\left(\frac{N+k}{N}\right)^{N+k}.$$

As

$$\begin{aligned}\log(m)&=k\bigl(\log(N)-\log(k)\bigr)+(N+k)\bigl(\log(N+k)-\log(N)\bigr)\\&\le k\log(N)+2k\end{aligned}$$

applying Theorem 4.6 yields the final bound. □

4.3. *Classic occupancy*

As mentioned in the Introduction, we can obtain bounds on (1.1) for the classic occupancy problem. Distribute m balls independently and uniformly among n urns. Let ξ_i be the number of balls in urn i. Then, $\xi_i\sim\mathrm{Bi}(m,n^{-1})$ and $\sum_i\xi_i=m$, and therefore

$$X_i=\frac{\xi_i-mn^{-1}}{\sqrt{m(1-n^{-1})}}$$

satisfies (2.9) and (3.1) and in addition $\sum_i\sigma_i^2=1$.

Theorem 4.7. *Let X be as above and let $Z\sim\mathrm{MVN}_n(0,\Sigma)$. Then, for any three times partially differentiable function h*,

$$\left|\mathbb{E}h(X)-\mathbb{E}h(Z)\right|\le\bigl(|h|_2+19|h|_3\bigr)\sqrt{\frac{n^2+4mn+6}{mn(n-1)}}.$$

Proof. We can apply (3.3), as X is exchangeable and (3.1) is satisfied. It is straightforward to verify that

$$\operatorname{Var}X_1^2=\frac{n^2+2(n-1)(m-3)}{n^2m(n-1)},\qquad \operatorname{Cov}\bigl(X_1^2,X_2^2\bigr)=-\frac{n^2-4n-2m+6}{mn^2(n-1)^2}$$

(see Lemma 4.8 below), which implies

$$\operatorname{Var}\sum_i X_i^2=n\operatorname{Var}X_1^2+\frac{n(n-1)}{2}\operatorname{Cov}\bigl(X_1^2,X_2^2\bigr)=\frac{n^2+4mn-2m-8n+6}{2mn(n-1)}.$$

Furthermore,

$$\mathbb{E}|X_1|^3\le\sqrt{\mathbb{E}X_1^2\mathbb{E}X_1^4}\le\sqrt{\frac{n^2+3nm-6n-3m+6}{mn^3(n-1)}}.$$

From this, the final bound follows. □

We record here some identities for the mixed moments of the ξ_i, which are easy to verify and needed in the above calculations.

Lemma 4.8. *Let ξ_1 and ξ_2 be the number of balls when distributing m balls uniformly and independently among n urns. Then*

$$\begin{aligned}
&\mathbb{E}\xi_1^2 = \frac{m}{n} + \frac{m(m-1)}{n^2},\\
&\mathbb{E}(\xi_1\xi_2) = \frac{m(m-1)}{n^2},\\
&\mathbb{E}\bigl(\xi_1^2\xi_2\bigr) = \frac{m(m-1)}{n^2} + \frac{m(m-1)(m-2)}{n^3},\\
&\mathbb{E}\bigl(\xi_1^2\xi_2^2\bigr) = \frac{m(m-1)}{n^2} + 2\frac{m(m-1)(m-2)}{n^3} + \frac{m(m-1)(m-2)(m-3)}{n^4}.
\end{aligned}$$

4.4. *Curie–Weiss model in the high-temperature regime*

Consider the n-spin system $\{-1,1\}^n$ with Hamiltonian

$$H(\sigma) = -\frac{1}{n}\sum_{i<j}\sigma_i\sigma_j.$$

To each configuration σ assign the probability

$$\mathbb{P}(\sigma) = \frac{\mathrm{e}^{-\beta H(\sigma)}}{Z(\beta)},$$

where $Z(\beta)$ is the normalising constant. This model is well-known as Curie–Weiss model; we refer to Eichelsbacher and Löwe [12] for a more detailed discussion of relevant literature. The authors of that article prove in particular bounds in univariate central limit theorems for the total magnetisation of this and similar models. Here, instead, we will estimate the error when we replace all the spins by corresponding Gaussian variables in the high-temperature regime $\beta < 1$; this, in particular, implies the central limit theorem for the total magnetisation.

Previous approaches using Stein's method to analyse the magnetisation of such models make use of exchangeable pairs (Eichelsbacher and Löwe [12] and Chatterjee and Shao [9]) which typically involves resampling a spin conditional on the other spins. It is worthwhile noting that the Stein coupling we will use does not require any resampling and, hence, does not form an exchangeable pair.

To avoid confusion with the notation σ_i for the spins, we will use s_{ij} instead of σ_{ij} to denote covariances in what follows.

Theorem 4.9. *Let $X_i = n^{-1/2}\sigma_i$ and let $Z \sim \mathrm{MVN}_n(0,\Sigma)$, where $\Sigma = (s_{ij})_{1\le i,j\le n}$ with*

$$s_{ij} = \begin{cases} \frac{1}{n} + \frac{\beta}{n^2(1-\beta)}, & \text{if } i = j,\\ \frac{\beta}{n^2(1-\beta)}, & \text{if } i \ne j. \end{cases}$$

Then, for $\beta < 1$,

$$\bigl|\mathbb{E}h(X) - \mathbb{E}h(Z)\bigr| \le C_\beta\left(\frac{|h|_1 + |h|_3}{n^{1/2}} + \frac{|h|_2}{n}\right)$$

for some constant C_β that only depends on β.

Proof. Define

$$m_i = \frac{1}{n}\sum_{j\neq i}\sigma_j, \qquad m = \frac{1}{n}\sum_i \sigma_i.$$

We recall the estimates

$$\mathbb{E}|m|^k \le C_\beta n^{-k/2};$$

see Eichelsbacher and Löwe [12], Lemma 3.5. Let (I, J, K, L) be distributed as in the proof of Theorem 3.3, independent of all else. Using the notation from Section 3.1 and the proof of Theorem 3.3 (with respect to exchangeability of X), define the vectors

$$X' = X^{(I)}, \qquad X'' = X^{IJKL}.$$

Set $\tilde{D} = D$ and define G as

$$G_k = -n^{3/2}\delta_{kK}\left(\frac{\beta}{n(1-\beta)} + \delta_{kI}\right)(\sigma_I - \beta m).$$

Define the matrix S as

$$S_{kl} = n^2\delta_{kK}\delta_{lI}s_{IK}.$$

Define now $f.$ as in the proof of Theorem 3.3. Then,

$$\begin{aligned}
&-\mathbb{E}\sum_k G_k f_k(X)\\
&\quad= n^{3/2}\mathbb{E}\left(\frac{\beta}{n(1-\beta)} + \delta_{KI}\right)(\sigma_I - \beta m) f_K(X)\\
&\quad= n^{-1/2}\mathbb{E}\sum_{i,k}\left(\frac{\beta}{n(1-\beta)} + \delta_{ki}\right)(\sigma_i - \beta m) f_k(X)\\
&\quad= n^{-1/2}\mathbb{E}\sum_{i,k}\left(\frac{\beta\sigma_i}{n(1-\beta)} - \frac{\beta^2 m}{n(1-\beta)} + \delta_{ki}\sigma_i - \delta_{ki}\beta m\right) f_k(X)\\
&\quad= n^{-1/2}\mathbb{E}\sum_k\left(\frac{\beta m}{1-\beta} - \frac{\beta^2 m}{1-\beta} + \sigma_k - \beta m\right) f_k(X)\\
&\quad= n^{-1/2}\mathbb{E}\sum_k \sigma_k f_k(X) = \mathbb{E}\sum_k X_k f_k(X)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\sum_k G_k f_k\bigl(X'\bigr)\\
&\quad= -n^{-3/2}\mathbb{E}\left(\frac{\beta}{n(1-\beta)} + \delta_{KI}\right)(\sigma_I - \beta m) f_K\bigl(X^{(I)}\bigr)\\
&\quad= -n^{-1/2}\mathbb{E}\sum_{i,k}\left(\frac{\beta}{n(1-\beta)} + \delta_{ki}\right)(\sigma_i - \beta m) f_k\bigl(X^{(i)}\bigr)\\
&\quad= -n^{-1/2}\mathbb{E}\sum_{i,k}\left(\frac{\beta}{n(1-\beta)} + \delta_{ki}\right)\bigl(\tanh(\beta m_i) - \beta m\bigr) f_k\bigl(X^{(i)}\bigr),
\end{aligned}$$

where for the last equation we used that $\mathbb{E}(\sigma_i|X^{(i)}) = \tanh(\beta m_i)$. Using the estimate

$$\begin{aligned}\left|\tanh(\beta m_i) - \beta m\right| &\le \left|\tanh(\beta m_i) - \tanh(\beta m)\right| + \left|\tanh(\beta m) - \beta m\right| \\ &\le \frac{\beta}{n} + \frac{\beta^3|m|^3}{6}\end{aligned}$$

we obtain

$$\left|\mathbb{E}\sum_k G_k f_k(X')\right| \le |f|_1 \frac{(6+\beta^2 n\mathbb{E}|m|^3)(\beta(1+\beta))}{6(1-\beta)n^{1/2}}$$

and hence

$$\left|\mathbb{E}R_0(t)\right| \le |h|_1 \frac{(6+\beta^2 n\mathbb{E}m^3)(\beta(1+\beta))}{6(1-\beta)n^{1/2}} \le \frac{C_\beta |h|_1}{n^{1/2}}$$

(cf. Remark 2.3). Using exchangeability of X for the second equation, the fact that $\delta_{KI} = \delta_{JL}$ and also that $s_{IK} = s_{JL}$, we obtain

$$\begin{aligned}&\mathbb{E}\sum_{k,l}(G_k D_l - S_{kl}) f_{kl}(X'') \\ &= n^2\mathbb{E}\left\{\left(\frac{1}{n}\left(\frac{\beta}{n(1-\beta)} + \delta_{KI}\right)(\sigma_I - \beta m)\sigma_I - s_{IK}\right) f_{KI}(X^{IJKL})\right\} \\ &= n^2\mathbb{E}\left\{\left(\frac{1}{n}\left(\frac{\beta}{n(1-\beta)} + \delta_{JL}\right)(\sigma_J - \beta m)\sigma_J - s_{JL}\right) f_{KI}(X)\right\} \\ &= \frac{1}{n}\mathbb{E}\sum_{i,j}\left\{\left(\frac{1}{n}\left(\frac{\beta}{n(1-\beta)} + 1\right)(\sigma_j - \beta m)\sigma_j - s_{jj}\right) f_{ii}(X)\right\} \\ &\quad + \frac{1}{n(n-1)}\mathbb{E}\sum_{i,j,k\neq i,l\neq j}\left\{\left(\frac{1}{n}\left(\frac{\beta}{n(1-\beta)}\right)(\sigma_j - \beta m)\sigma_j - s_{jl}\right) f_{ki}(X)\right\} \\ &= \mathbb{E}\left\{\sum_j\left(\frac{1}{n}\left(\frac{\beta}{n(1-\beta)} + 1\right)(1 - \beta\sigma_j m) - s_{jj}\right)\sum_i \frac{f_{ii}(X)}{n}\right\} \\ &\quad + \mathbb{E}\left\{\sum_{j,l\neq j}\left(\frac{1}{n}\left(\frac{\beta}{n(1-\beta)}\right)(1 - \beta\sigma_j m) - s_{jl}\right)\sum_{i,k\neq i}\frac{f_{ki}(X)}{n(n-1)}\right\} \\ &= -\mathbb{E}\left\{\frac{(\beta + n(1-\beta))\beta m^2}{n(1-\beta)}\sum_i \frac{f_{ii}(X)}{n} - \frac{(n-1)\beta^2 m^2}{n(1-\beta)}\sum_{i,k\neq i}\frac{f_{ki}(X)}{n(n-1)}\right\}.\end{aligned}$$

Thus,

$$\left|\mathbb{E}R_1(t)\right| \le C_\beta |h|_2 \mathbb{E}m^2 \le \frac{C_\beta |h|_2}{n}.$$

Now

$$\begin{aligned}\mathbb{E}R_2(t,u) &= \mathbb{E}\sum_{k,l,m}(G_k D_l - S_{kl}) D'_m f_{klm}(X + uD') \\ &= n^{3/2}\mathbb{E}\sum_{m\in\{I,J,K,L\}}\left[\frac{1}{n}\left(\frac{\beta}{n(1-\beta)} + \delta_{KI}\right)(\sigma_I - \beta m)\sigma_I - s_{IK}\right]\sigma_m f_{KIm}(X + uD').\end{aligned}$$

From this, it is not difficult to see that

$$\left|\mathbb{E}R_2(t,u)\right| \le \frac{C_\beta |h|_3}{n^{1/2}}$$

(recall the definition of s_{ij} and note that the probability that $I = K$ is $1/n$). Similarly,

$$\left|\mathbb{E}R_3(t,u)\right| \le \frac{C_\beta |h|_3}{n^{1/2}}.$$

Putting all the estimates together, yields the claim. □

Acknowledgements

The author thanks the anonymous referees for helpful comments and literature references.

References

[1] J. Baik and T. M. Suidan. A GUE central limit theorem and universality of directed first and last passage site percolation. *Int. Math. Res. Not.* **2005** (2005) 325–337. MR2131383
[2] A. D. Barbour, M. Karoński and A. Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *J. Combin. Theory Ser. B* **47** (1989) 125–145. MR1047781
[3] T. Bodineau and J. Martin. A universality property for last-passage percolation paths close to the axis. *Electron. Commun. Probab.* **10** (2005) 105–112 (electronic). MR2150699
[4] E. Bolthausen. Exact convergence rates in some martingale central limit theorems. *Ann. Probab.* **10** (1982) 672–688. MR0659537
[5] P. Carmona and Y. Hu. Universality in Sherrington–Kirkpatrick's spin glass model. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** (2006) 215–222. MR2199799
[6] S. Chatterjee. A simple invariance theorem. Preprint, 2005. Available at http://arxiv.org/abs/math.PR/0508213.
[7] S. Chatterjee. A generalization of the Lindeberg principle. *Ann. Probab.* **34** (2006) 2061–2076. MR2294976
[8] S. Chatterjee and E. Meckes. Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** (2008) 257–283. MR2453473
[9] S. Chatterjee and Q.-M. Shao. Non-normal approximation by Stein's method of exchangeable pairs with application to the Curie–Weiss model. *Ann. Appl. Probab.* **21** (2011) 464–483. MR2807964
[10] L. H. Y. Chen and A. Röllin. Stein couplings for normal approximation. Preprint, 2010. Available at http://arxiv.org/abs/1003.6039.
[11] A. Dembo and Y. Rinott. Some examples of normal approximations by Stein's method. In *Random Discrete Structures (Minneapolis, MN, 1993)* 25–44. *IMA Vol. Math. Appl.* **76**. Springer, New York, 1996. MR1395606
[12] P. Eichelsbacher and M. Löwe. Stein's method for dependent random variables occurring in statistical mechanics. *Electron. J. Probab.* **15** (2010) 962–988. MR2659754
[13] L. Erdős, H.-T. Yau and J. Yin. Bulk universality for generalized Wigner matrices. Preprint, 2010. Available at arxiv.org/abs/1001.3453.
[14] F. Götze and A. N. Tikhomirov. Limit theorems for spectra of random matrices with martingale structure. *Teor. Veroyatn. Primen.* **51** (2006) 171–192. MR2324173
[15] I. G. Grama. On moderate deviations for martingales. *Ann. Probab.* **25** (1997) 152–183. MR1428504
[16] K. Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** (2000) 437–476. MR1737991
[17] K. Johansson. Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Comm. Math. Phys.* **215** (2001) 683–705. MR1810949
[18] E. S. Meckes. On Stein's method for multivariate normal approximation. In *High Dimensional Probability V: The Luminy Volume* 153–178. Inst. Math. Statist., Beachwood, OH, 2009. MR2797946
[19] E. Mossel, R. O'Donnell and K. Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Ann. of Math.* **171** (2010) 295–341. MR2630040
[20] M. Raič. A multivariate CLT for decomposable random vectors with finite second moments. *J. Theoret. Probab.* **17** (2004) 573–603.
[21] G. Reinert and A. Röllin. Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** (2009) 2150–2173. MR2573554
[22] I. Rinott and V. I. Rotar'. Some estimates for the rate of convergence in the CLT for martingales. II. *Theory Probab. Appl.* **44** (1999) 523–536. MR1805821
[23] Y. Rinott and V. Rotar'. A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Anal.* **56** (1996) 333–350. MR1379533
[24] V. I. Rotar'. Certain limit theorems for polynomials of degree two. *Teor. Veroyatn. Primen.* **18** (1973) 527–534. Actual title is "Some limit theorems for polynomials of degree two." MR0326803
[25] D. Slepian. The one-sided barrier problem for Gaussian noise. *Bell System Tech. J.* **41** (1962) 463–501. MR0133183

[26] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory* 583–602. Univ. California Press, Berkeley, CA, 1972. MR0402873

[27] T. Suidan. A remark on a theorem of Chatterjee and last passage percolation. *J. Phys. A* **39** (2006) 8977–8981. MR2240468

[28] M. Talagrand. The Parisi formula. *Ann. of Math.* **163** (2006) 221–263. MR2195134

[29] M. Talagrand. *Mean Field Models for Spin Glasses. Volume I*. Springer-Verlag, Berlin, 2010. MR2731561

[30] T. Tao and V. Vu. Random matrices: universality of local eigenvalue statistics. *Acta Math.* **206** (2011) 127–204. MR2784665