Stein's method for functional approximations

Mikołaj J. Kasprzak

St. Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2018

Abstract

We extend the ideas of Barbour's paper from 1990 and adapt Stein's method for distributional approximation to infinite-dimensional distributions. Hence, we obtain theoretical results bounding the rate of functional convergence of certain classes of stochastic processes to diffusions. Those are applied to examples coming from queuing theory, random-graph theory, statistics and combinatorics.

We firstly look at the motivation for this thesis and an overview of Stein's method. Then we present our original work, contained in four articles. The first one is a collaboration with Andrew Duncan and Sebastian Vollmer, published in the *Electronic Communications in Probability* and the other ones, for which I am the sole author, are currently under consideration for publication. The first paper corrects a mistake in Barbour's seminal work from 1990.

The second paper considers the approximation of a time-changed Poisson process by a time-changed Brownian motion for time changes independent of the processes they are applied to. As an application, we study the M/M/1 queue and a time-changed Brownian Motion and bound a distance between the two.

The third paper studies the asymptotic behaviour of scaled sums of random vectors having different dependence structures. As an application, a bound on the distance between scaled non-degenerate U-statistics and Brownian Motion is proved. Moreover, we prove a quantitative functional limit theorem for exceedances in the m-scans process.

In the fourth paper, we adapt the exchangeable-pair approach to Stein's method to approximations by infinite-dimensional laws. It is used to provide the rate of convergence in a functional combinatorial central limit theorem, extending the result of Barbour and Janson from 2009. We further apply this approach to study the asymptotics of edge and two-star counts in a certain graph-valued process.

The final part of the thesis presents the conclusions and suggestions for future work.

Acknowledgements

I would like to thank Professor Gesine Reinert for suggesting the topic of this thesis, for her exceptional guidance and support, many very interesting mathematical discussions and her direction on my way to becoming an independent researcher. It has been a true pleasure to be her doctoral student.

I am also grateful to all the other people whom I have been lucky to discuss my research with. I would particularly like to thank Andrew Barbour, Christian Döbler, Arnaud Doucet, Alison Etheridge, Christina Goldschmidt, Jonathan Huggins, Lester Mackey, Ivan Nourdin, Giovanni Peccati, Dino Sejdinovic and Yvik Swan. I am confident that those discussions were not only a source of new research ideas but also helped me become a better mathematician overall.

My time here in Oxford would have not been so enjoyable and productive if it had not been for all the fantastic people from the Statistics Department I have had the pleasure of meeting and talking to every day. I am particularly grateful to the faculty members who I have not already mentioned: Julien Berestycki, Paul Chleboun, James Martin, David Steinsaltz, and Matthias Winkel; my fellow DPhil students: Andreas Anastasiou, James Ayre, Serte Donderwinkel, Mitch Gooding, Eleonora Kreacic, Ian Letter, Vlad Margarint, Sarah Penington, Emilka Pompe, Dane Rogers, Franz Rembart, David Rindt, Freddy Sørensen, Aaron Smith, Romek Stasiński, Andi Wang, Dominic Yeo and many more that there is not a space to thank.

Finally, I would like to thank all those I have met through the Oxford University Polish Society for supporting me in maintaining a healthy work-life balance and encouraging me to take a break from my DPhil research every now and then. Some of the events organised by the Society will be my fondest memories of the time I spent in Oxford; I am confident that some of the friendships that began there will last long after I leave the University.

My research has been funded by an EPSRC Studenship.

Contents

Part I: Opening remarks

1 Introduction

In this thesis, we will study the problem of obtaining functional approximations, together with bounds on the corresponding rates of convergence, using Stein's method. We will consider the framework of [Bar90] and study distances between distributions of random objects X and Y taking values in the Skorokhod space $D([0,1],\mathbb{R}^d)$. Specifically, we will provide bounds on quantities of the form $|\mathbb{E}g(X) - \mathbb{E}g(Y)|$ for test functions $g : D([0,1],\mathbb{R}^d) \to \mathbb{R}$ which are twice Fréchet differentiable and whose second Fréchet derivative is Lipschitz. Depending on the particular example looked at, we will occasionally consider classes of test functions restricted to those elements whose first and second Fréchet derivatives are bounded. In most cases, nevertheless, the test functions we use will form convergence-determining classes, in the sense that weak-convergence results will follow from our bounds as corollaries. The aim of my DPhil has been to prove general results which would be applicable to a wide range of examples, coming, for instance, from queuing theory or the theory of random graphs.

1.1 Motivation

My current and planned future research does not only fill a theoretical gap in the literature but has also direct relevance to applications. Researchers studying real-life discrete phenomena often choose to model them with scaling limits of discrete processes rather than those processes themselves. This approach allows them to look at the phenomena of interest "from a distance" so that the models they use become more robust to changes in the local details and easier to study using stochastic analysis. For instance, researchers in population genetics often choose to model the behaviour of a large finite population with the Wright-Fisher diffusion rather than Markov chains, such as the Moran model or the discrete Wright-Fisher model. Hence, they implicitly assume the population size to be infinite. Similarly, premium-claim money flows in risk science are often modelled by Bessel processes rather than Bessel-like random walks converging to them.

Nevertheless, due to the lack of theoretical results measuring the quality of those approximations, it is often impossible to determine how accurate the conclusions driven from the analysis of the limiting diffusions are and how well they describe the real-world phenomena of interest. My research using Stein's method, alongside the work of [CD13], opens the door for developing bounds on the rate of convergence in those approximations, hence increasing their applicability. This door is also opened by some of Decreusefond's more recent work, for instance [BDM18]. While [CD13, BDM18] and some other papers of Decreusefond and coauthors concentrate on approximations on a carefully chosen Hilbert space, my results are derived for functionals acting on the much richer Skorohod space of càdlàg paths.

1.2 Contents of the thesis

Attached are four articles: [KDV17], a collaboration with Andrew Duncan and Sebastian Vollmer, and [Kas17a, Kas17b, Kas18], for which I am the sole author. The first one of those is published in the *Electronic Communications in Probability*, while the latter three are under consideration for publication. Here is a brief description of those papers.

[KDV17] corrects a mistake in [Bar90] and therefore makes it possible to extend the results of [Bar90], which is achieved in the other papers constituting this thesis. In [Bar90] Barbour creates a framework for Stein's method to be used for infinite-dimensional distributions and provides a bound on the rate of convergence in Donsker's functional central limit theorem. He also studies scaled sums of locally dependent one-dimensional random variables and their distance from Brownian Motion. Paper [Kas17a] considers the problem of providing an explicit bound on the quality of the approximation of a compensated Poisson process by a Wiener process and, via time changes, extends it to diffusion approximations of a certain class of continuous-time Markov chains. This result is applied to an example coming from queuing theory. Paper [Kas18] provides a framework for obtaining functional approximations of sums of random variables valued in $\mathbb{R}^d$ with different local dependence structures and applies it to examples related to U-statistics and random graphs. Finally, [Kas17b] builds on [Kas18] by applying the so-called "exchangeable-pair approach" in the functional setting to globally weakly dependent structures.

This thesis is structured as follows. Section 2 of this document provides an introduction to Stein's method including its version catering for approximations by infinite-dimensional distributions. Section 3 outlines the research I completed during my DPhil, which this thesis is based on. It is followed by the four papers constituting the main part of the thesis. The last part of this document contains a discussion of the results, conclusions, as well as suggestions for future work and plans.

2 Stein's method for distributional approximation

2.1 Overview of Stein's method

In his seminal paper [Ste72], Charles Stein introduced a method for proving normal approximations and obtained an explicit bound on the rate of convergence to the standard normal distribution. Suppose that the aim is to approximate the expectation $\mathbb{E}h(W)$, for some random variable W, by the expectation $\mathbb{E}h(Z)$ for Z standard normal. Stein observed that a random variable Z has standard normal law if and only if $\mathbb{E}Zf(Z) = \mathbb{E}f'(Z)$ for all smooth functions f. Therefore, if, for a random variable W with mean 0 and variance 1, $\mathbb{E}f'(W) - \mathbb{E}Wf(W)$ is close to zero for a large class of functions f, then the law of W should be approximately Gaussian. He then proposed that, instead of evaluating $|\mathbb{E}h(W) - \mathbb{E}h(Z)|$ directly for a given function h, one can first find an $f = f_h$ solving the following *Stein equation*:

$$f'(w) - wf(w) = h(w) - \mathbb{E}h(Z)$$

and then find a bound on $|\mathbb{E}f'(W) - \mathbb{E}Wf(W)|$. This approach often turns out to be much easier than a direct evaluation of $|\mathbb{E}h(W) - \mathbb{E}h(Z)|$, due to some bounds on the solutions f_h, which can be derived in terms of the derivatives of h. In many situations it also turns out to be more powerful than estimating the speed of weak convergence using Lévy's criterion related to convergence of characteristic functions. Indeed, characteristic functions are extremely useful in the asymptotic study of sums of independent random variables yet fail to adapt well in the presence of dependence. Numerous examples showing how well Stein's method handles local (and in some cases also global) dependence of random variables whose sums are of interest, are presented, for instance, in [BC05], as well as in my work [Kas18], [Kas17b].

Since the seminal work of Stein was published, his method has been significantly developed and extended to approximations by distributions other than normal. The paper [Che75] on the Poisson approximation gave rise to the well-know Chen-Stein method. Generalising the approach used for the normal distribution, the aim of Stein's method is to find a bound of the quantity $|\mathbb{E}_{\nu_n} h - \mathbb{E}_\mu h|$, where μ is the target (known) distribution, ν_n is the approximating law and h is chosen from a suitable class of real-valued test functions $\mathcal{H}$. For instance, considering measures on the real line and taking $\mathcal{H} = \{\mathbb{I}_{(-\infty,a]} : a \in \mathbb{R}\}$ would let us obtain the supremum distance between the cumulative distribution functions corresponding to the two laws μ and ν_n, known as the Kolmogorov-Smirnov distance. In general, proving the existence of bounds converging to 0 could serve as a proof of weak convergence as long as the class of functions $\mathcal{H}$ is a convergence determining class, i.e.

$$\left(\forall h \in \mathcal{H} \quad \mathbb{E}_{\nu_n} h \xrightarrow{n\to\infty} \mathbb{E}_\mu h\right) \quad \Longrightarrow \quad \nu_n \xrightarrow{W} \mu,$$

where $\xrightarrow{W}$ denotes weak convergence. The idea of Stein's method is to find an operator (called *Stein operator*) $\mathcal{A}$ acting on a class of real-valued functions such that

$$\big(\forall f \in \text{Domain}(\mathcal{A}) \quad \mathbb{E}_\nu \mathcal{A} f = 0\big) \quad \Longleftrightarrow \quad \nu = \mu,$$

where μ is our target distribution. In the next step, for a given function $h \in \mathcal{H}$, a solution $f = f_h$ to the following *Stein equation*

$$\mathcal{A} f = h - \mathbb{E}_\mu h$$

is sought and its properties studied. Finally, using various mathematical tools (among which the most popular are Taylor expansions in the continuous case, Malliavin calculus and coupling methods), a bound is to be found for the quantity $|\mathbb{E}_{\nu_n} \mathcal{A} f_h|$. This approach often turns out to be much easier than trying to find a bound on $|\mathbb{E}_{\nu_n}h - \mathbb{E}_\mu h|$ directly, in particular in the presence of dependence.

To find a suitable Stein operator, Barbour [Bar88] and Götze [Göt91] developed the so-called *generator approach*, which made it possible to adapt the method to many other probability laws. They noticed that the generator of a Markov Process with a stationary law μ may be used as a Stein operator for approximation by μ. An example is the generator of the Ornstein-Uhlenbeck diffusion which may be used as the Stein operator for normal approximation. Stein's method for distributions including gamma (e.g. [Luk94]) and general univariate and some multivariate distributions has also been developed (see [LRS17], [MRS18]). Furthermore, a powerful connection has been found between Stein's method and Malliavin calculus which has led to many new results concerning primarily normal approximations (see [NP12]).

Furthemore, the exchangeable-pair approach of [Ste86] proves to be particularly useful in presence of local or weak global dependence. This method was used to derive bounds on the distance between a univariate random variable W and a standard normal random variable Z via constructing a random variable W' such that (W, W') is an exchangeable pair and the linear regression condition

$$\mathbb{E}\left[W' - W|W\right] = -\lambda W$$

is satisfied for some $\lambda > 0$. In [RR97], an additional error term R is allowed to occur so that the condition now takes the form

$$\mathbb{E}[W' - W|W] = -\lambda W + R$$

and the following bound is derived

$$\sup_x \left|\mathbb{P}[W \leq x] - \mathbb{P}[Z \leq x]\right| \leq \frac{6}{\lambda}\sqrt{\mathrm{Var}\mathbb{E}[(W'-W)^2|W]} + \frac{6}{\lambda^{1/2}}\sqrt{\mathbb{E}|W'-W|^3} + \frac{19}{\lambda}\sqrt{\mathrm{Var}R}.$$

This method has been extended to non-normal univariate laws, for instance in [CDM05], [Röl07], and to multivariate laws, in [CM08], [RR09], [Mec09], [RR10]. In a piece of work constituting part of this thesis, [Kas17b], the method is extended to multivariate functional approximations.

An accessible account of the method can be found in the surveys [LRS17] and [Ros11] as well as the books [BHJ92] and [CGS11], which treat the cases of Poisson and normal approximation, respectively, in detail. A database of information and publications connected to Stein's method is in [Swa16].

2.2 Stein's method for infinite-dimensional distributions

Approximations by infinite-dimensional laws have not been covered in Stein's method literature very widely, with the notable exceptions of [Bar90], [BJ09] and recently [CD13] and [BDM18]. They are nevertheless an interesting topic as there are a large class of functional-limit results available in the literature and used in applications but without explicit bounds on the quality of the approximation.

An example of such a class of results is the theory of Stroock and Varadhan [SV79]. As presented in [Dur96], the theory describes the scaling-limit behaviour of a sequence of Markov chains Y^h indexed by a parameter $h > 0$, taking values in a set $E_h \subset \mathbb{R}^d$ and satisfying certain assumptions. Specifically, for the transition kernel Π_h of Y^h and the corresponding scaled transition kernel

$$K_h(x, dy) = \frac{1}{h}\Pi_h(x, dy),$$

the following quantities are defined for $1 \leq i, j \leq d$

$$\begin{aligned} a^h_{i,j}(x) &= \int_{|y-x|\leq 1} (y_i - x_i)(y_j - x_j)K_h(x, dy), \\ b^h_i(x) &= \int_{|y-x|\leq 1} (y_i - x_i)K_h(x, dy), \\ \Delta^h_\epsilon(x) &= K_h\left(x, B(x, \epsilon)^c\right). \end{aligned}$$

The coefficients $a_{i,j}$ and b_i are assumed to be continuous functions on $\mathbb{R}^d$. Further, it is assumed that the martingale problem $M(a, b)$ below is well-posed, that is there exists a unique (up to distribution) process $(X_t, 0 \le t \le 1)$, such that $X_0 = x$ almost surely, and

$$M_t^i = X_t^i - \int_0^t b_i(X_s)ds \quad \text{and} \quad M_t^i M_t^j - \int_0^t a_{i,j}(X_s)ds$$

are local martingales. Stroock and Varadhan proved that, if, additionally, for each $1 \le i, j \le d$ and for every $R > 0, \epsilon > 0$,

$$\lim_{h \to 0} \sup_{|x| \le R} \left| a_{ij}^h(x) - a_{ij}(x) \right| = 0,$$

$$\lim_{h \to 0} \sup_{|x| \le R} \left| b_i^h(x) - b_i(x) \right| = 0,$$

$$\lim_{h \to 0} \sup_{|x| \le R} \Delta_\epsilon^h(x) = 0$$

and $X_t^h = Y_{\lfloor t/h \rfloor}^h$ for $t \in [0,1]$ then $(X_t^h, t \in [0,1])$ converges weakly to $(X_t, t \in [0,1])$ in the Skorokhod topology, where X_t solves the martingale problem $M(a,b)$. Similar results hold for continuous-time Markov chains. There is no universal method of providing the rate of convergence in those results.

A special case of the results of Stroock and Varadhan is the well-known Donsker's theorem stating that a properly rescaled random walk converges in law to a Wiener process. In order to obtain a quantitative version of Donsker's theorem, in [Bar90], Barbour constructs a Markov Process

$$W(t,u) = \sum_{k \ge 0} X_k(u) S_k(t); \quad 0 \le t \le 1, u \ge 0,$$

where $\{X_k\}_{k \ge 0}$ is a collection of independent, identically distributed Ornstein-Uhlenbeck processes on $[0, \infty)$, with equilibrium distribution $\mathcal{N}(0,1)$, and S_k are the Schauder functions. Those are defined by

$$S_0(t) = t; S_k(t) = \int_0^t H_k(u)du, k \ge 1,$$

where, for $2^n \le k < 2^{n+1}$,

$$H_k(u) = 2^{n/2}\{\mathbb{1}[2^{-n}k - 1 \le u \le 2^{-n}(k+1/2) - 1] - \mathbb{1}[2^{-n}(k+1/2) - 1 \le u \le 2^{-n}(k+1) - 1]\}.$$

In [Bar90], Barbour then notes that the stationary distribution for the process $\big(W(\cdot,u)\big)_{u\geq 0}$ is the Wiener measure and constructs the infinitesimal generator $\mathcal{A}$ of this process, given by:

$$\mathcal{A}f(w) = -Df(w)[w] + \sum_{k\geq 0} D^2 f(w)[S_k^{(2)}]. \tag{1}$$

The domain of the generator is taken to be the set of twice differentiable functions $g : D([0,1],\mathbb{R}) \to \mathbb{R}$, such that:

$$\begin{aligned}\|g\|_M := \sup_{w\in D([0,1],\mathbb{R})} \frac{|g(w)|}{1+\|w\|_\infty^3} + \sup_{w\in D([0,1],\mathbb{R})} \frac{\|Dg(w)\|}{1+\|w\|_\infty^2} + \sup_{w\in D([0,1],\mathbb{R})} \frac{\|D^2g(w)\|}{1+\|w\|_\infty} \\ + \sup_{w,h\in D([0,1],\mathbb{R})} \frac{\|D^2g(w+h)-D^2g(w)\|}{\|h\|_\infty} < \infty, \end{aligned} \tag{2}$$

where $D^k g$ denotes the k-th Frechet derivative of g, and $\|D^k g\| = \sup_{\|h\|=1} \left|D^k g[h,h,\cdots,h]\right|$. Using the general theory of Markov Processes and their stationary laws, the author concludes that, for any f in the domain of $\mathcal{A}$ and B denoting the Wiener process on the time interval $[0,1]$,

$$\mathbb{E}\mathcal{A}f(B) = 0.$$

Therefore, $\mathcal{A}$ is a Stein operator for approximation by the distribution of a Wiener process indexed by time in $[0,1]$.

Barbour then solves the corresponding Stein equation and uses Taylor expansions together with smoothness properties of the solution to obtain bounds on the distance between a scaled random walk

$$Y(t) = n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} \tilde{X}_i, t \in [0,1] \tag{3}$$

and a Wiener process $(B(t), 0 \leq t \leq 1)$, where $(\tilde{X}_i)_{i=1}^n$ are real random variables with mean zero, unit variance and finite third absolute moment. Specifically, it is proved that for a constant C

$$|\mathbb{E}g(Y) - \mathbb{E}g(B)| \leq C\|g\|_M n^{-1/2}(\mathbb{E}|\tilde{X}_1|^3 + \sqrt{\log n})$$

for every function $g : D([0,1],\mathbb{R}) \to \mathbb{R}$ satisfying condition (2). The result is obtained in two steps, the first one bounding the distance between Y and a scaled Gaussian random walk and the second one concerning the distance between the Gaussian random walk and Brownian Motion. The term of order $n^{-1/2}\sqrt{\log n}$ comes from the Brownian modulus of continuity invoked in the second step. It may be avoided if one compares the continuous approximands, obtained by linearly interpolating between the points $\left(j/n, Y(j/n)\right)_{j=0}^{n}$, to Brownian Motion. This kind of two-step approximation will be a repeating theme in the proofs of the results included in this thesis. [Bar90] also presents a number of applications including adding short-range dependence between the summands in the scaled sum (3) and considering dissociated random summands.

Reference [BJ09] establishes a functional version of the Hoeffding combinatorial central limit theorem. It considers a sequence of real matrices $a^{(n)} := (a_0^{(n)}(i,j), 1 \leq i,j \leq n), n \geq 1$ satisfying the property $\sum_{j=1}^{n} a(i,j) = 0$ for all i, and a random permutation π of the set $\{1, 2, \cdots, n\}$. The process

$$Y(t) = \frac{1}{s(a)} \sum_{i=1}^{\lfloor nt \rfloor} a(i, \pi(i))$$

is then studied, where $s(a)$ is an appropriate scaling factor, making $\text{Var}[Y(1)] = 1$. Barbour and Janson first use Stein's method to establish a bound on a convergence-determining distance (i.e. one which metrizes weak convergence) of Y from the pre-limiting Gaussian process

$$Z_n = \sum_{i=1}^{\lfloor nt \rfloor} \tilde{Z}_i,$$

where the $\tilde{Z}_i$'s are centred Gaussian with the same covariance structure as the $\frac{1}{s(a)} a(i, \pi(i))$'s. They then establish convergence of the pre-limiting process Z_n to a continuous Gaussian process Z, given by

$$Z(t) = \sigma_a^{-1} \int_{[0,t]\times[0,1]} \alpha(v,w) K(dv, dw).$$

In the expression above, K is the Kiefer process $K = W(v,w) - vW(1,w)$ for W denoting a two-dimensional Brownian sheet, α is the L_2-limit of functions $\alpha_n(v,w) := a^{(n)}(\lceil nv \rceil, \lceil nw \rceil)$ and $\sigma_a := \|\alpha\|_{L_2} = \left(\int_0^1 \int_0^1 \left(\alpha(u,v)\right)^2 du dv\right)^{1/2}$. Barbour and Janson thus prove convergence of

Y_n to the continuous Gaussian process Z. They then establish a bound on a non-convergence-determining distance of the pre-limiting process from the continuous Gaussian process.

The last two papers [CD13] and [BDM18], in this brief survey, focus on Brownian approximations in subspaces of the Skorokhod space, called the Besov-Liouville spaces, equipped with the L^2 topology. We will not present the definition of the spaces due to its lengthiness and technicality but we note that a full exposition can be found in [SKM93].

In particular, in [CD13], Coutin and Decreusefond use Malliavin calculus to develop an abstract version of Stein's method for Hilbert valued random variables and provide quantitative results related to Poisson approximations of Brownian Motion, Donsker's theorem and the linear interpolation of Brownian Motion. An example of the result they prove is a bound on the rate of convergence of a scaled sum of Rademacher random variables, with discontinuities removed via linear interpolations, to Brownian Motion. The rate depends on the parameters of the Besov-Liouville space considered but may be made of order $n^{-1/2}$ on a carefully chosen space.

3 Summary of completed research

3.1 A note on A. Barbour's paper on Stein's method for diffusion approximations

In [Bar90], Barbour uses the generator approach to Stein's method in order to obtain a bound on the distance between a scaled random walk and Brownian Motion. Having found an operator $\mathcal{A}$, in (1), acting on a class of real-valued functions taking arguments in the Skorokhod space, such that:

$$\mathbb{E}_\mu \mathcal{A} f = 0 \quad \forall f \in \text{Domain}(\mathcal{A}) \quad \iff \quad \mu \text{ is the Wiener measure,}$$

Barbour considers the following Stein equation:

$$\mathcal{A}f = g - \mathbb{E}g(B), \tag{4}$$

where B is a Brownian Motion. Since $\mathcal{A}$ is interpreted as the generator of a Markov process whose stationary law is the one of B, the semigroup $(T_t)_{t\geq 0}$ of this process is found and using [EK86, Proposition 1.5, page 9], for a fixed g a solution to (4) is represented in the following way:

$$f = \phi(g) = -\int_0^\infty T_u g du. \tag{5}$$

However, [EK86, Proposition 1.5, page 9] requires strong continuity of the semigroup $(T_t)_{t\geq 0}$, which, as noted by Andrew Duncan and Sebastian Vollmer, does not hold in the setup considered in this case. Having spoken to Sebastian Vollmer and having heard about his idea for a counterexample showing the violation of strong continuity, I made the counterexample rigorous and followed the proof of [EK86, Proposition 1.5, page 9] in order to show that (5) nevertheless solves (4) for a certain class of functions g and that the main assertions of [Bar90] still hold. The results have been published in [KDV17].

3.2 Diffusion approximations via Stein's method and time changes

Reference [Kas17a] is the first step towards finding a method of estimating the rate of convergence in the limit theorems comprising the Stroock-Varadhan theory of diffusion approximation [SV79].

In [Kas17a] I obtained bounds on the distance between a sum of compensated Poisson processes, with time changes independent of those processes applied to them, and a time-changed Brownian Motion. As described in [EK86, Section 4, Chapter 6], continuous-time Markov chains whose jump sizes belong to a countable set may be represented as sums of time-changed Poisson processes, each of which takes account of the jumps in a given direction. Specifically, I have proved the following:

Theorem 3.1. *Let $X_1, X_2, ...$ be i.i.d. with mean 0, variance 1 and finite third moment and*

$s:[0,1]\to[0,\infty)$ be a strictly increasing, continuous function with $s(0)=0$. Define:

$$Y_n(t)=n^{-1/2}\sum_{i=1}^{\lfloor ns(t)\rfloor}X_i,\quad t\in[0,1]$$

and let $(Z(t),t\in[0,1])=\big(B(s(t)),t\in[0,1]\big)$, where B is a standard Brownian Motion. Suppose that $\|g\|_M<\infty$, as given by (2). Then

$$\begin{aligned}|\mathbb{E}g(Y_n)-\mathbb{E}g(Z)|\leq&\|g\|_M\frac{30+54\cdot 5^{1/3}s(1)}{\sqrt{\pi}\log 2}n^{-1/2}\sqrt{\log(2s(1)n)}\\&+\|g\|_M s(1)\left(1+\left(\frac{3}{2}\right)^3\sqrt{\frac{2}{\pi}}s(1)^{3/2}\right)\mathbb{E}|X_1|^3n^{-1/2}\\&+\|g\|_M\frac{2160}{\sqrt{\pi}(\log 2)^{3/2}}n^{-3/2}(\log(2s(1)n))^{3/2}.\end{aligned}$$

I have also shown the following result:

Theorem 3.2. *Suppose that P is a Poisson process with rate 1 and $S^{(n)}:[0,1]\to[0,\infty)$ is a sequence of increasing deterministic continuous functions, such that $S^{(n)}(0)=0$. Let $S:[0,1]\to[0,\infty)$ be also increasing and continuous. Let $Z(t)=B(S(t)),t\in[0,1]$ where B is a standard Brownian Motion and*

$$\tilde{Y}_n(t)=\frac{P\left(nS^{(n)}(t)\right)-nS^{(n)}(t)}{\sqrt{n}},\qquad t\in[0,1].$$

Then, for all g with $\|g\|_M<\infty$, as given by (2):

$$\begin{aligned}|\mathbb{E}g(\tilde{Y}_n)-\mathbb{E}g(Z)|\leq\|g\|_M&\left\{\left(2+\frac{27\sqrt{2}}{2\sqrt{\pi}}S(1)\right)\sqrt{\|S-S^{(n)}\|}+\frac{27\sqrt{2}}{2\sqrt{\pi}}\|S-S^{(n)}\|^{3/2}\right.\\&+n^{-1/2}\left[\frac{30+54\cdot 5^{1/3}S(1)}{\sqrt{\pi}\log 2}\sqrt{\log(2S(1)n)}\right.\\&+\left.\left(1+\left(\frac{3}{2}\right)^3\sqrt{\frac{2}{\pi}}S^{(n)}(1)^{3/2}\right)S^{(n)}(1)(1+2e^{-1})+1+\frac{\log(1+2e^{-1})+2\log n}{\log\log(n+2)}\right]\\&+n^{-1}\frac{9\sqrt{S^{(n)}(1)}}{2}\left(1+3nS^{(n)}(1)\right)^{1/2}\left[4+\frac{16701+128(\log n)^3}{(\log\log(n+3))^3}\right]^{1/3}\end{aligned}$$

$$+ n^{-3/2}\left[\frac{2160}{\sqrt{\pi}(\log 2)^{3/2}}(\log(2S(1)n))^{3/2} + 8 + \frac{33402 + 256(\log n)^3}{(\log\log(n+3))^3}\right]\Bigg\}.$$

Using these results, I found a bound on the distance between a rescaled M/M/1 queue and a time-changed Brownian Motion. Here is the setup, which is based on [Rob03]. For two independent rate 1 Poisson processes $\mathbf{P}_1$ and $\mathbf{P}_{-1}$ the M/M/1 queue with jump rates λ and μ is the solution to the following equation:

$$\mathbf{L}(t) = \mathbf{L}(0) + \mathbf{P}_1(\lambda t) - \int_0^t \mathbb{1}_{\{\mathbf{L}(s-)>0\}}\mathbf{P}_{-1}(\mu ds). \tag{6}$$

Let $(x_n)_{n\geq 1}$ and x be such that $\frac{x_n}{n} \xrightarrow{n\to\infty} x \in \mathbb{R}_+$. Consider the renormalised proces

$$\overline{\mathbf{L}_n}(t) = \frac{\mathbf{L}(nt)}{n}, \quad t \in [0,1], \tag{7}$$

where $(\mathbf{L}(t), t \in [0,1])$ solves (6) with initial condition x_n. The first part of [Rob03, Proposition 5.16] states that $\overline{\mathbf{L}_n}$ of (7) converges with respect to uniform topology to the function $t \mapsto \overline{\mathbf{L}}(t) := \big(x + (\lambda - \mu)t\big)^+$. The second part of the Proposition states that, if $\sqrt{n}(\frac{x_n}{n} - x) \xrightarrow{n\to\infty} 0$ then the process

$$\mathbf{Y}_n(t) = \sqrt{n}\left(\overline{\mathbf{L}_n}(t) - \overline{\mathbf{L}}(t)\right), \quad t \in [0,1]$$

converges in distribution to $\Big(\mathbf{B}\left((\lambda+\mu)t\right), t \in [0,1]\Big)$, where $\mathbf{B}$ is a standard Brownian Motion. I proved a theorem which provides a bound on the rate of this convergence with respect to a convergence-determining class of tests functions. It is worth noting that a similar bound is obtained in [BDM18] yet with respect to a restricted class of test functions which does not metrize weak convergence in the Skorokhod topology.

3.3 Functional approximations of multivariate sums

The paper [Kas18] extends the results of [Bar90] to approximations of multi-dimensional scaled sums of (possibly) dependent random variables by Gaussian processes. Specifically, [Kas18] considers processes of the form:

$$\mathbf{Y}_n(t) = \left(\sum_{i=1}^{\lambda_1} X_{i,1} J_{i,1}(t), \cdots, \sum_{i=1}^{\lambda_p} X_{i,p} J_{i,p}(t) \right), \quad t \in [0,1],$$

where $\lambda_j \leq n$ for all j, the X_i's are locally dependent (there exists a set $\mathbb{A}_i \subset \{1, \cdots, n\}$ such that X_i are independent of $\{X_j : j \in \mathbb{A}_i^c\}$), and the functions $J_{i,k}$ are independent of the Z_i. The paper provides a bound on the distance of Y_n from $Z = \Sigma^{1/2}$, where B is a standard p-dimensional Brownian Motion and Σ is a positive-definite covariance matrix. In this paper the results hold for functions g such that $\|g\|_{M^1} < \infty$, where

$$\begin{aligned} \|g\|_{M^1} := \sup_{w \in D^p} \frac{|g(w)|}{1 + \|w\|^3} + \sup_{w \in D^p} \|Dg(w)\| + \sup_{w \in D^p} \|D^2 g(w)\| \\ + \sup_{w,h \in D^p} \frac{\|D^2 f(w+h) - D^2 f(w)\|}{\|h\|} < \infty. \end{aligned}$$

This class of funtions is smaller than the previously considered class of functions g satisfying $\|g\|_M < \infty$, as in (2).

The paper proves that for any g with $\|g\|_{M^1} < \infty$ we have

$$|\mathbb{E} g(\mathbf{Y}_n) - \mathbb{E} g(Z)| \leq \sum_{i=1}^{7} \epsilon_i,$$

where

$$\epsilon_1 = \frac{1}{6} \sum_{i=1}^{n} \mathbb{E} \left\{ \left(\sum_{k,l,m=1}^{p} \left[\left(X_{i,k}\right)^2 \|J_{i,k}\|^2 \mathbb{1}_{[1,\lambda_k]}(i) \left(\sum_{j \in \mathbb{A}_i} X_{j,l} \|J_{j,l}\| \mathbb{1}_{[1,\lambda_l]}(j) \right)^2 \right.\right.\right.$$

$$\cdot \left(\sum_{j \in \mathbb{A}_i} X_{j,m} \|J_{j,m}\| \mathbb{1}_{[1,\lambda_m]}(j) \right)^2 \Bigg] \Bigg) \Bigg)^{1/2} \Bigg\} ;$$

$$\epsilon_2 = \frac{1}{3} \sum_{i=1}^{n} \sum_{j \in \mathbb{A}_i} \sum_{k,l=1}^{p} \mathbb{E} \Bigg\{ \Bigg[\sum_{m=1}^{p} \Bigg(X_{i,k} \|J_{i,k}\| X_{j,l} \|J_{j,l}\| \mathbb{1}_{[1,\lambda_k]}(i) \mathbb{1}_{[1,\lambda_l]}(j)$$

$$\cdot \sum_{r \in \mathbb{A}_j \cap \mathbb{A}_i^c} X_{r,m} \|J_{r,m,n}\| \mathbb{1}_{[1,\lambda_m]}(r) \Bigg)^2 \Bigg]^{1/2} \Bigg\} ;$$

$$\epsilon_3 = \frac{1}{3} \sum_{i=1}^{n} \sum_{j \in \mathbb{A}_i} \sum_{k,l=1}^{p} \Bigg\{ \mathbb{E} \left[\left| X_{i,k} X_{j,l} \right| \right] \mathbb{1}_{[1,\lambda_k]}(i) \mathbb{1}_{[1,\lambda_l]}(j)$$

$$\cdot \mathbb{E} \left[\|J_{i,k}\| \, \|J_{j,l}\| \sqrt{ \sum_{m=1}^{p} \left(\sum_{r \in \mathbb{A}_i \cup \mathbb{A}_j} X_{r,m} \|J_{r,m}\| \mathbb{1}_{[1,\lambda_m]}(r) \right)^2 } \right] \Bigg\} ;$$

$$\epsilon_4 = \frac{K}{2} \sum_{i=1}^{n} \sum_{k,l=1}^{p} \left| \frac{\Sigma_{k,l}}{\sqrt{\lambda_k \lambda_l}} - \mathbb{E}[X_{i,k} X_{i,l}] \right| ;$$

$$\epsilon_5 = \frac{K}{2} \sum_{i=1}^{n} \sum_{j \in \mathbb{A}_i \setminus \{i\}} \sum_{k,l=1}^{p} \left| \mathbb{E}[X_{i,k} X_{j,l}] \right| ;$$

$$\epsilon_6 = \frac{6\sqrt{5}}{\sqrt{2 \log 2}} \left(\sum_{i=1}^{p} \frac{\log(2\lambda_i)}{\lambda_i} \right)^{1/2} \left(\sum_{i=1}^{p} \Sigma_{i,i} \right)^{1/2} ;$$

$$\epsilon_7 = \sum_{k=1}^{p} \sum_{i=1}^{\lambda_k} \frac{\sqrt{\Sigma_{ii}}}{\sqrt{\lambda_k}} \mathbb{E} \left\| J_{i,k} - \mathbb{1}_{[i/\lambda_k, 1]} \right\| .$$

If the summands are independent, i.e. $\mathbb{A}_i = \{i\}$ for all i, then ϵ_2 and ϵ_5 disappear from the bound and ϵ_1 and ϵ_3 become simpler. The new bound takes the following form

$$\left| \mathbb{E} g(\mathbf{Y}_n) - \mathbb{E} g(\mathbf{Z}) \right| \leq \|g\|_{M^1} \left(\epsilon_1 + \epsilon_3 + \epsilon_4 + \epsilon_6 + \epsilon_7 \right),$$

where:

$$\epsilon_1 = \frac{1}{6}\sum_{i=1}^{n}\mathbb{E}\left\{\left[\sum_{k,l,m=1}^{p}\left(X_{i,k}X_{i,l}X_{i,m}\|J_{i,k}\|\,\|J_{i,l}\|\,\|J_{i,m}\|\mathbb{1}_{[1,\lambda_k]\cap[1,\lambda_l]\cap[1,\lambda_m]}(i)\right)^2\right]^{1/2}\right\};$$

$$\epsilon_3 = \frac{1}{3}\sum_{k,l=1}^{p}\sum_{i=1}^{\min(\lambda_k,\lambda_l)}\left\{\mathbb{E}\left[\left|X_{i,k}X_{i,l}\right|\right]\mathbb{E}\left[\|J_{i,k}\|\,\|J_{i,l}\|\sqrt{\sum_{m=1}^{p}\left(X_{i,m}\|J_{i,m}\|\mathbb{1}_{[1,\lambda_m]}(i)\right)^2}\right]\right\};$$

$$\epsilon_4 = \frac{K}{2}\sum_{i=1}^{n}\sum_{k,l=1}^{p}\left|\frac{\Sigma_{k,l}}{\sqrt{\lambda_k\lambda_l}} - \mathbb{E}[X_{i,k}X_{i,l}]\right|;$$

$$\epsilon_6 = \frac{6\sqrt{5}}{\sqrt{2\log 2}}\left(\sum_{i=1}^{p}\frac{\log(2\lambda_i)}{\lambda_i}\right)^{1/2}\left(\sum_{i=1}^{p}\Sigma_{i,i}\right)^{1/2};$$

$$\epsilon_7 = \sum_{k=1}^{p}\sum_{i=1}^{\lambda_k}\frac{\sqrt{\Sigma_{ii}}}{\sqrt{\lambda_k}}\mathbb{E}\left\|J_{i,k,n} - \mathbb{1}_{[i/\lambda_k,1]}\right\|.$$

Terms $\epsilon_1, \epsilon_2, \epsilon_3$ correspond to a Berry-Esseen-type bound involving third moments of the summands, and also account for local dependence between the summands. Terms ϵ_4 and ϵ_5 involve a variance estimation with the latter corresponding to the off-diagonal terms of the covariance matrix of the summands, accounting for the dependence. Term ϵ_6 comes from estimates on the moments of the Brownian modulus of continuity and accounts for the transition from the Skorokhod space to the Wiener space of continuous functions. Term ϵ_7 describes the randomness of the functions $J_{i,k}$ and their distance from indicators $\mathbb{1}_{[i/\lambda_k,1]}$.

This result is used to find the rate of convergence of scaled bivariate non-degenerate U-statistics to Brownian Motion. It is also applied in order to prove a quantitative functional limit result for exceedances in an m-scans process. Specifically, I consider an extension of the one-dimensional [CGS11, Example 9.2, p. 254] to the multidimensional and functional setting. For $j = 1, 2, \cdots$, I let $V_j = \left(V_{j,1}, \cdots, V_{j,p}\right)$ be i.i.d. random vectors in $\mathbb{R}^p$. Furthermore, for $k = 1, \cdots, p$ and $i = 1, \cdots, n$, I let $R_{i,k} = \sum_{l=0}^{m-1} V_{i+l,k}$ be an m-scans process. I take $a = (a_1, \cdots, a_p) \in \mathbb{R}^p$ and suppose that $n > m$.

For $k = 1, \cdots, p$, I let $\pi_k = \mathbb{P}(R_{1,k} \leq a_k)$ and for $i = 1, \cdots, n$ and $k = 1, \cdots, p$,

$$X_{i,k} = \frac{1}{n}\left(\sum_{j=1}^{n} \mathbb{1}[R_{n(i-1)+j,k} \leq a_k]\right) - \pi_k.$$

Moreover, I use the following notation $X_i = (X_{i,1}, \cdots, X_{i,p})$ for $i = 1, \cdots, n$. I consider

$$\mathbf{Y}_n(t) = \sum_{i=1}^{\lfloor nt \rfloor} \left(X_{i,1}, \cdots, X_{i,p}\right) \quad t \in [0,1]$$

and let $\Sigma \in \mathbb{R}^{p\times p}$ be given by

$$\Sigma_{k,l} = \psi_{k,l}(0) + \sum_{d=1}^{m-1} \left(\psi_{l,k}(d) + \psi_{k,l}(d)\right).$$

Finally, I bound the distance between $\mathbf{Y}_n$ and $\mathbf{Z} = \Sigma^{1/2}\mathbf{B}$, where $\mathbf{B}$ is a standard p-dimensional Brownian Motion.

3.4 Functional approximations via exchangeable pairs

The project [Kas17b] extends the multivariate exchangeable-pair approach to Stein's method of [RR09] to the functional setting. Specifically, [Kas17b] proves an abstract approximation theorem. It considers an exchangeable pair of stochastic processes (Y_n, Y_n'), taking values in the Skorokhod space $D([0,1], \mathbb{R}^p)$, and satisfying the following linear regression condition

$$Df(Y_n)[Y_n] = 2\mathbb{E}^{Y_n} Df(Y_n)\left[(Y_n - Y_n')\Lambda_n\right] + R_f, \tag{8}$$

where $\mathbb{E}^{Y_n}[\cdot] := \mathbb{E}\left[\cdot|Y_n\right]$, for all f with $\|f\|_M < \infty$, as defined in (2), some $\Lambda_n \in \mathbb{R}^{p\times p}$ and some random variable $R_f = R_f(Y_n)$. It states that the distance between Y_n and a scaled sum of Gaussian random vectors D_n can be bounded by a quantity depending on the distance between Y_n and Y_n' and the covariance structure of $Y_n - Y_n'$ and D_n:

$$\left|\mathbb{E}g(Y_n) - \mathbb{E}g(D_n)\right| \leq R_1 + R_2 + R_3,$$

for any $g \in M$ as defined in (2), where f is the g-solution of the corresponding Stein equation and

$$
\begin{aligned}
R_1 &= \frac{\|g\|_M}{6}\mathbb{E}\|(Y_n - Y_n')\Lambda_n\|\|Y_n - Y_n'\|^2,\\
R_2 &= \left|\mathbb{E}D^2 f(Y_n)\left[(Y_n - Y_n')\Lambda_n, Y_n - Y_n'\right] - \mathbb{E}D^2 f(Y_n)\left[D_n, D_n\right]\right|,\\
R_3 &= |\mathbb{E}R_f|.
\end{aligned}
$$

The role of Λ_n in condition (8) is equivalent to that played by Λ^{-1} in [RR09] for Λ defined by (1.7) therein. Indeed, [RR09] considers an exchangeable pair (W, W') of centred $\mathbb{R}^d$-valued random vectors satisfying condition

$$\mathbb{E}^W(W' - W) = -\Lambda W + R \tag{9}$$

for an invertible matrix Λ and some $R = R(W)$ and derives a bound on the distance of W from a Gaussian vector. Condition (8) is, however, more appropriate in the functional setting than a straightforward adaptation of the condition (9) of [RR09]. This is due to the fact that for general processes Y_n the properties of the Frechet derivative do not allow us to treat evaluating the derivative in the direction of $Y_n - Y_n'$ as matrix multiplication and multiplying both sides of the hypothetical condition

$$-Df(Y_n)[\Lambda Y_n] = \mathbb{E}^{Y_n} Df(Y_n)[Y_n - Y_n']$$

by Λ^{-1} does not recover an expression for $-Df(Y_n)[Y_n]$, as desired in the Stein operator (1) or similar Stein operators.

The abstract approximation theorem of [Kas17b] is then applied to prove a functional combinatorial central limit theorem. It considers a scaled sum of elements $X_{i,\pi(i)}$ of an array $\mathbb{X} = \{X_{i,j} : i, j = 1, \cdots, n\}$ of independent random variables, where π is a random permutation on $\{1, \cdots, n\}$. It states that, under certain assumptions, this scaled sum converges to a continuous Gaussian process with a known covariance structure. Furthermore, it provides a bound on its distance from a certain pre-limiting Gaussian-mixture process. This is an improvement of similar results of [BJ09] which concern deterministic arrays $\mathbb{X}$. In a further application, [Kas17b] provides an explicit bound on the rate of convergence to a two-dimensional Gaussian process of the joint distribution of rescaled edge and two-star counts in a Bernoulli-graph-valued process. The bound is of order $\frac{\sqrt{\log n}}{\sqrt{n}}$, similar to the bound Barbour [Bar90] obtained for the Brownian approximation of a random walk. In our case, however, the class of test functions is restricted to those real-valued functions g acting on the Skorokhod space, which satisfy

$$\sup_{w\in D^p}\frac{|g(w)|}{1+\|w\|^3}+\sup_{w\in D^p}\frac{\|Dg(w)\|}{1+\|w\|}+\sup_{w\in D^p}\frac{\|D^2g(w)\|}{1+\|w\|}+\sup_{w,h\in D^p}\frac{\|D^2f(w+h)-D^2f(w)\|}{\|h\|}<\infty,$$

where $\|\cdot\|=\|\cdot\|_\infty$. This class of test functions determines convergence in distribution, as shown in [BJ09, Proposition 3.1]. As in the majority of results presented in this thesis, the bounds in those two applications are achieved in a two-step process: through a pre-limiting approximation by a piecewise constant Gaussian process and an approximation of that one by a continuous Gaussian process.

Part II: One article and three preprints

Electron. Commun. Probab. **22** (2017), no. 23, 1–8.
DOI: 10.1214/17-ECP54
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Note on A. Barbour's paper on Stein's method for diffusion approximations

Mikołaj J. Kasprzak* Andrew B. Duncan† Sebastian J. Vollmer‡

Abstract

In [2] foundations for diffusion approximation via Stein's method are laid. This paper has been cited more than 130 times and is a cornerstone in the area of Stein's method (see, for example, its use in [1] or [7]). A semigroup argument is used in [2] to solve a Stein equation for Gaussian diffusion approximation. We prove that, contrary to the claim in [2], the semigroup considered therein is not strongly continuous on the Banach space of continuous, real-valued functions on $D[0,1]$ growing slower than a cubic, equipped with an appropriate norm. We also provide a proof of the exact formulation of the solution to the Stein equation of interest, which does not require the aforementioned strong continuity. This shows that the main results of [2] hold true.

Keywords: Stein's method ; Donsker's Theorem ; Diffusion approximations.
AMS MSC 2010: Primary: 60B10, 60F17, Secondary: 60J60, 60J65, 60E05.
Submitted to ECP on February 22, 2017, final version accepted on April 4, 2017.

1 Introduction

In [2] a claim is made that the semigroup defined by (2.4) thereof is strongly continuous on space L defined on page 299 thereof. We prove that this is not the case. Nevertheless, we show that the only assertion of the paper following from the aforementioned assumption of strong continuity, namely the claim that (2.20) solves the Stein equation (2.1), remains true. This may be proved by adapting the proof of [5, Proposition 9, p. 9] and noting that in the case of interest in [2], the point-wise continuity of the semigroup is sufficient. It then follows that all the other results of [2] hold true.

In Section 2 we recall the relevant definitions and notation from [2]. In Section 3 we give a counterexample to the strong continuity of the semigroup. In Section 4 we provide a proof of the fact that the function (2.20) of [2] does actually solve the Stein equation. We do this by following the steps of the proof of [5, Proposition 9, p. 9] and proving each of the assertions therein for the semigroup of interest by hand.

2 Definitions and notation

By $D = D[0,1]$ we will mean the Skorohod space of all the càdlàg functions $w : [0,1] \to \mathbb{R}$. In the sequel $\|\cdot\|$ will always denote the supremum norm. By $D^k f$ we mean the k-th Fréchet derivative of f and the k-linear norm B is defined to be $\|B\| =$

*University of Oxford, United Kingdom. E-mail: kasprzak@stats.ox.ac.uk
†University of Sussex, United Kingdom. E-mail: andrew.duncan@sussex.ac.uk
‡University of Warwick, United Kingdom. E-mail: svollmer@turing.ac.uk

$\sup_{\{h:\|h\|=1\}}|B[h,...,h]|$. We will also often write $D^2f(w)[h^{(2)}]$ instead of $D^2f(w)[h,h]$. Let:

$$L=\left\{f:D\to\mathbb{R}:f\text{ is continuous and }\sup_{w\in D}\frac{|f(w)|}{1+\|w\|^3}<\infty\right\}$$

and for any $f\in L$ let $\|f\|_L=\sup_{w\in D}\frac{|f(w)|}{1+\|w\|^3}$.

We define:

$$\|f\|_M=\sup_{w\in D}\frac{|f(w)|}{1+\|w\|^3}+\sup_{w\in D}\frac{\|Df(w)\|}{1+\|w\|^2}+\sup_{w\in D}\frac{\|D^2f(w)\|}{1+\|w\|}+\sup_{w,h\in D}\frac{\|D^2f(w+h)-D^2f(w)\|}{h}$$

for any $f\in L$ for which the expressions exist and

$$M=\left\{f\in L:f\text{ is twice Fréchet differentiable and }\|f\|_M<\infty\right\}.$$

The Stein operator for approximation by Z, the Brownian Motion on $[0,1]$, is defined, as in (2.9) and (2.11) of [2], by:

$$\mathcal{A}f(w)=-Df(w)[w]+\mathbb{E}D^2f(w)\left[Z^{(2)}\right]=-Df(w)[w]+\sum_{k\geq 0}D^2f(w)\left[S_k^{(2)}\right],$$

for any $f:D[0,1]\to\mathbb{R}$, for which it exists. By $(S_k)_{k\geq 0}$ we denote the Schauder functions defined, as on page 299 of [2] by:

$$S_0(t)=t;\qquad S_k(t)=\int_0^t H_k(u)du,\quad k\geq 1,$$

where, for $2^n\leq k<2^{n+1}$:

$$H_k(u)=2^{n/2}\left(\mathbb{1}\left[\frac{k}{2^n}-1\leq u\leq\frac{k+\frac{1}{2}}{2^n}-1\right]-\mathbb{1}\left[\frac{k+\frac{1}{2}}{2^n}-1<u\leq\frac{k+1}{2^n}-1\right]\right).$$

We also define a semigroup acting on L:

$$(T_uf)(w)=\mathbb{E}\left[f\left(we^{-u}+\sigma(u)Z\right)\right],\tag{2.1}$$

where $\sigma^2(u)=1-e^{-2u}$.

For any $g\in M$ with $\mathbb{E}g(Z)=0$, the Stein equation is given by:

$$\mathcal{A}f=g.$$

The idea of Stein's method applied in [2] is to find a bound on $\mathbb{E}\mathcal{A}f(X)$, where f is a solution to this equation, in order bound $|\mathbb{E}g(X)-\mathbb{E}g(Z)|$, for some stochastic process X on $[0,1]$.

3 Counterexample to strong continuity

It is well known that the Ornstein-Uhlenbeck semigroup is not strongly continuous on the space $C_b(\mathbb{R})$, see [3]. More generally, given a separable Hilbert space H, in [8] it is noted that this semigroup is also not strongly continuous on the space $C_{b,k}$ of all continuous functions $\psi:H\to\mathbb{R}$ such that $x\to\psi(x)/(1+|x|^k)$ is uniformly continuous and $\sup_{x\in H}\frac{\psi(x)}{1+|x|^k}<\infty$. Following these two results, in this section we shall show that the semigroup T_u defined by (2.1) is not strongly continuous on the Banach space L by constructing an explicit counterexample.

Lemma 3.1. *The semigroup T_u is not strongly continuous on $(L,\|\cdot\|_L)$.*

Proof. Consider $f \in L$ defined by:

$$f(w) = (1 + \|w\|^3)\sin(\|w\|).$$

Note that:

$$\begin{aligned}
\|T_u f - f\|_L &= \sup_{w\in D} \frac{\left|\mathbb{E}(1+\|we^{-u}+\sigma(u)Z\|^3)\sin(\|we^{-u}+\sigma(u)Z\|) - (1+\|w\|^3)\sin(\|w\|)\right|}{1+\|w\|^3} \\
&= \sup_{w\in D} \Big|\mathbb{E}\sin(\|we^{-u}+\sigma(u)Z\|) - \sin(\|w\|) \\
&\quad + \frac{\mathbb{E}\left[\left(\|we^{-u}+\sigma(u)Z\|^3 - \|w\|^3\right)\sin(\|we^{-u}+\sigma(u)Z\|)\right]}{1+\|w\|^3}\Big| \\
&\geq \sup_{w\in D} \left|\mathbb{E}\sin(\|we^{-u}+\sigma(u)Z\|) - \sin(\|w\|)\right| \\
&\quad - \sup_{w\in D}\left|\frac{\mathbb{E}\left[\left(\|we^{-u}+\sigma(u)Z\|^3 - \|w\|^3\right)\sin(\|we^{-u}+\sigma(u)Z\|)\right]}{1+\|w\|^3}\right| \\
&\geq \sup_{w\in D}\left|\sin(e^{-u}\|w\|) - \sin(\|w\|)\right| - \sup_{w\in D}\left|\mathbb{E}\sin(\|we^{-u}+\sigma(u)Z\|) - \sin(e^{-u}\|w\|)\right| \\
&\quad - \sup_{w\in D}\left|\frac{\mathbb{E}\left[\left(\|we^{-u}+\sigma(u)Z\|^3 - \|w\|^3\right)\sin(\|we^{-u}+\sigma(u)Z\|)\right]}{1+\|w\|^3}\right|. \qquad (3.1)
\end{aligned}$$

Now:

$$\begin{aligned}
&\sup_{w\in D}\left|\frac{\mathbb{E}\left[\left(\|we^{-u}+\sigma(u)Z\|^3 - \|w\|^3\right)\sin(\|we^{-u}+\sigma(u)Z\|)\right]}{1+\|w\|^3}\right| \\
\leq &\sup_{w\in D}\frac{\mathbb{E}\left|\left(\|we^{-u}+\sigma(u)Z\| - \|w\|\right)\left(\|we^{-u}+\sigma(u)Z\|^2 + \|we^{-u}+\sigma(u)Z\|\|w\| + \|w\|^2\right)\right|}{1+\|w\|^3} \\
\leq &\sup_{w\in D}\frac{\mathbb{E}\left[\left(\|w\|(1-e^{-u}) + \sigma(u)\|Z\|\right)\left(\|w\|^2(2e^{-2u}+e^{-u}+1) + \sigma(u)\|Z\|\|w\| + 2\sigma^2(u)\|Z\|^2\right)\right]}{1+\|w\|^3} \\
= &\sup_{w\in D}\frac{1}{1+\|w\|^3}\cdot\Big\{\|w\|^3(1-e^{-u})(2e^{-2u}+e^{-u}+1) + \|w\|^2\mathbb{E}\|Z\|\sigma(u)\left[2e^{-2u}+2\right] \\
&+\|w\|\sigma^2(u)\mathbb{E}\|Z\|^2\left[2(1-e^{-u})+1\right] + 2\sigma^3(u)\mathbb{E}\|Z\|^3\Big\} \xrightarrow{u\searrow 0} 0. \qquad (3.2)
\end{aligned}$$

Furthermore, given $\epsilon > 0$, consider $R > 0$ such that $\mathbb{P}(\|Z\| > R) < \epsilon$. Fix $\delta > 0$, such that for any $a, b \in \mathbb{R}$: $|a-b| < \delta \Rightarrow |\sin(a) - \sin(b)| < \epsilon$. Now, for any u such that $\sigma(u)R < \delta$ and for every $w \in D$, we have:

$$\|Z\| \leq R \Longrightarrow \left|\|we^{-u}+\sigma(u)Z\| - e^{-u}\|w\|\right| \leq \sigma(u)\|Z\| < \delta$$

and so:

$$\begin{aligned}
&\left|\mathbb{E}\sin(\|we^{-u}+\sigma(u)Z\|) - \sin(e^{-u}\|w\|)\right| \\
\leq &\mathbb{E}\left|\sin(\|we^{-u}+\sigma(u)Z\|) - \sin(e^{-u}\|w\|)\right|\mathbb{1}\left[\|Z\|\leq R\right] \\
&+ \mathbb{E}\left|\sin(\|we^{-u}+\sigma(u)Z\|) - \sin(e^{-u}\|w\|)\right|\mathbb{1}\left[\|Z\| > R\right] \\
\leq &\epsilon + 2\epsilon.
\end{aligned}$$

Therefore:

$$\sup_{w\in D}\left|\mathbb{E}\sin(\|we^{-u}+\sigma(u)Z\|) - \sin(e^{-u}\|w\|)\right| \xrightarrow{u\searrow 0} 0. \qquad (3.3)$$

Finally, for any $k \in \mathbb{N}$, consider $w_k \in D$ defined by $w_k(t) = k\pi$. For $u_k = -\log\left(1 - \frac{1}{2k}\right) \xrightarrow{k\to\infty} 0$, we have:

$$\left|\sin(e^{-u_k}\|w\|) - \sin(\|w\|)\right| = \left|\sin\left(k\pi - \frac{\pi}{2}\right) - \sin(k\pi)\right| = 1.$$

Therefore:

$$\exists (u_k)_{k=1}^{\infty}: \quad u_k \xrightarrow{k\to\infty} 0 \quad \text{and} \quad \sup_{w\in D}\left|\sin(e^{-u_k}\|w\|) - \sin(\|w\|)\right| \geq 1. \tag{3.4}$$

By (3.1), (3.2), (3.3), (3.4), $\lim_{u\to 0}\|T_u f - f\|_L \neq 0$ and so T_u is not strongly continuous on $(L, \|\cdot\|_L)$. □

4 Solution to the Stein equation

We first show that the function, which in Lemma 4.3 is shown to solve the Stein equation, exists and belongs to the domain of $\mathcal{A}$.

Lemma 4.1. *For any $g \in M$, such that $\mathbb{E}[g(Z)] = 0$, $f = \phi(g) = -\int_0^\infty T_u g du$ exists and is in the domain of $\mathcal{A}$.*

Proof. Note that:

$$|g(w) - g(x)| \leq C_g(1 + \|w\|^2 + \|x\|^2)\|w - x\| \tag{4.1}$$

uniformly in $w, x \in D[0,1]$. This follows from the fact that:

$$\begin{aligned}
&|g(w) - g(x)| \leq \|g\|_M\|w - x\|^3 + \left|Dg(x)[w - x] + \frac{1}{2}D^2g(x)[w - x, w - x]\right| \\
\leq&\|g\|_M\|w - x\|^3 + \|Dg(x)\|\|w - x\| + \frac{1}{2}\|D^2g(x)\|\|w - x\|^2 \\
\leq&\|g\|_M\|w - x\|\left(\|w - x\|^2 + 1 + \|x\|^2 + \frac{1}{2}\|w - x\|(1 + \|x\|)\right) \\
\leq&\|g\|_M\|w - x\|\left(2\|w\|^2 + 2\|x\|^2 + 1 + \|x\|^2 + \frac{1}{2}(\|w\| + \|x\| + \|w\|\|x\| + \|x\|^2)\right) \\
\leq&C_g(1 + \|w\|^2 + \|x\|^2)\|w - x\|
\end{aligned}$$

uniformly in w, x because $\|w\| \leq 1 + \|w\|^2$, $\|x\| \leq 1 + \|x\|^2$ and $\|w\|\|x\| \leq \|w\|^2 + \|x\|^2$. Now, we note that, as a consequence of (4.1), we have:

$$\begin{aligned}
&\lim_{t\to\infty}\int_0^t |T_u g(w)|\, du = \lim_{t\to\infty}\int_0^t \left|\mathbb{E}g(we^{-u} + \sigma(u)Z)\right| du \\
\leq&\lim_{t\to\infty}\left[\int_0^t \left|\mathbb{E}\left[g(we^{-u} + \sigma(u)Z) - g(\sigma(u)Z)\right]\right| du + \int_0^t |\mathbb{E}[g(\sigma(u)Z) - g(Z)]|\, du\right] \\
\leq&C_g\lim_{t\to\infty}\left[\int_0^t \mathbb{E}\left[\left(1 + \|e^{-u}w + \sigma(u)Z\|^2 + \sigma^2(u)\|Z\|^2\right)e^{-u}\|w\|\right] du\right. \\
&\left.+ \int_0^t \mathbb{E}\left|(1 + (\sigma^2(u) + 1)\|Z\|^2)\right| \|(\sigma(u) - 1)Z\|\, du\right] \\
\leq&C_g\lim_{t\to\infty}\left[\int_0^t \left[e^{-u}\|w\| + 2e^{-3u}\|w\|^3 + 3\sigma^2(u)e^{-u}\|w\|\mathbb{E}\|Z\|^2\right] du\right. \\
&\left.+ \int_0^t (\sigma(u) - 1)\mathbb{E}\left|(1 + (\sigma^2(u) + 1)\|Z\|^2)\right| \|Z\|\, du\right] \\
\leq&C(1 + \|w\|^3),
\end{aligned} \tag{4.2}$$

for some constant C. Since L is complete, this guarantees the existence of $\phi(g)$.

As noted in (2.23) and (2.24) of [2], dominated convergence may be used, because of (4.2) to obtain that:

$$D^k\phi(g)(w) = -\int_0^\infty e^{-ku}D^k g(we^{-u}+\sigma(u)Z)du, \quad k=1,2. \tag{4.3}$$

and, as a consequence, that $\phi(g)\in M$. This is enough to conclude that $\phi(g)$ belongs to the domain of $\mathcal{A}$ by the observation directly above the formulation of $\mathcal{A}$ labelled as (2.9) in [2]. □

Remark 4.2. The argument of (2.23) and (2.24) in [2] also readily gives that for any $g\in M$ and $t>0$: $\int_0^t T_u g du \in M$.

We now prove that observation (2.19) of [2] is true for all $g\in M$:

Lemma 4.3. *For all $t>0$ and for all $g\in M$:*

$$T_t g - g = \mathcal{A}\left(\int_0^t T_u g du\right). \tag{4.4}$$

Proof. We will follow the steps of the proof of Proposition 1.5 on p. 9 of [5]. Observe that for all $w\in D[0,1]$ and $h>0$:

$$\begin{aligned}
&\frac{1}{h}[T_h - I]\int_0^t T_u g(w)du = \frac{1}{h}\int_0^t [T_{u+h}g(w) - T_u g(w)]du \\
=&\frac{1}{h}\int_t^{t+h} T_u g(w)du - \frac{1}{h}\int_0^h T_u g(w)du \\
\overset{(2.1)}{=}&\frac{1}{h}\int_t^{t+h}\mathbb{E}[g(we^{-u}+\sigma(u)Z)]du - \frac{1}{h}\int_0^h \mathbb{E}[g(we^{-u}+\sigma(u)Z)]du.
\end{aligned} \tag{4.5}$$

Taking $h\to 0$ on the left-hand side gives $\mathcal{A}\left(\int_0^t T_u g(w)du\right)$, since $\int_0^t T_u g(w)du$ belongs to the domain of $\mathcal{A}$ by Lemma 4.1 and Remark 4.2. In order to analyse the right-hand side note that:

$$\begin{aligned}
&\left|\frac{1}{h}\int_0^h \mathbb{E}[g(we^{-u}+\sigma(u)Z)] - g(w)du\right| \\
\overset{\text{MVT}}{\le}&\frac{1}{h}\int_0^h \mathbb{E}\left[\|w(e^{-u}-1)+\sigma(u)Z\| \sup_{c\in[0,1]}\|Dg\left(cw+(1-c)(we^{-u}+\sigma(u)Z)\right)\|\right]du \\
\le&\frac{\|g\|_M}{h}\int_0^h \mathbb{E}\left[\left(\|w\|(1-e^{-u})+\sigma(u)\|Z\|\right)\left(1+3\|w\|^2+3\|w\|^2e^{-2u}+3\sigma^2(u)\|Z\|^2\right)\right]du \\
=&\frac{\|g\|_M}{h}\mathbb{E}\Big\{\left(1+3\|w\|^2+3\|Z\|^2\right)(\|w\|(-1+h+\cosh(h)-\sinh(h)) \\
&+\|Z\|e^{-h}(-\sqrt{e^{2h}-1}+e^h(h+\log(1+e^{-h}\sqrt{-1+e^{2h}}))\Big) \\
&+3\|w\|(\|w\|^2-\|Z\|^2)\left(\frac{e^{-3h}}{6}(e^h-1)^2(e^h+2)\right) \\
&+3(\|w\|^2\|Z\|-\|Z\|^3)\frac{1}{3}\left(\sqrt{1-e^{-2h}}-\sqrt{e^{-6h}(e^{2h}-1)}\right)\Big\} \xrightarrow{h\to 0} 0.
\end{aligned} \tag{4.6}$$

Similarly:

$$\left|\frac{1}{h}\int_t^{t+h}\mathbb{E}\left[g(we^{-u}+\sigma(u)Z)\right]du - \mathbb{E}\left[g(we^{-t}+\sigma(t)Z)\right]\right| \xrightarrow{h\to 0} 0.$$

Therefore, as $h\to 0$, the right-hand side of (4.5) converges to $T_t g - g$, which finishes the proof. □

Proposition 4.4. *For any $g \in M$, such that $\mathbb{E}g(Z) = 0$, $f = \phi(g) = -\int_0^\infty T_u g du$ solves the Stein equation:*

$$\mathcal{A}f = g.$$

Proof. We note that for any $h > 0$ and for any $f \in M$:

$$\frac{1}{h}\left[T_{n,s+h}f - T_s f\right] = T_s\left[\frac{T_h - I}{h}f\right].$$

We also note that for any $w \in D[0,1]$, $g \in M$ and some constant K_1 depending only on f:

$$\left|T_u f(w) - f(w) - \mathbb{E}Df(w)\left[\sigma(u)Z - w(1-e^{-u})\right] - \frac{1}{2}\mathbb{E}D^2 f(w)\left[\left\{\sigma(u)Z - w(1-e^{-u})\right\}^{(2)}\right]\right| \\ \leq K_1(1+\|w\|^3)u^{3/2}, \tag{4.7}$$

as noted on page 300 of [2]. Therefore, we can apply dominated convergence to obtain:

$$\begin{aligned}\left(\frac{d}{ds}\right)^+ T_s f(w) &= \lim_{h\searrow 0} T_s\left[\frac{T_h - I}{h}f(w)\right] = \lim_{h\searrow 0}\mathbb{E}\left[\frac{T_h - I}{h}f(we^{-s} + \sigma(s)Z)\right] \\ &= \mathbb{E}\left[\lim_{h\searrow 0}\frac{T_h - I}{h}f(we^{-s} + \sigma(s)Z)\right] = T_s\mathcal{A}f(w).\end{aligned}$$

Similarly, for $s > 0$, $\left(\frac{d}{ds}\right)^- T_s f = T_s\mathcal{A}f$ because:

$$\begin{aligned}&\lim_{h\searrow 0}\frac{1}{-h}\left[T_{s-h}f - T_s f\right](w) - T_s\mathcal{A}f(w) \\ &= \lim_{h\searrow 0} T_{s-h}\left[\left(\frac{T_h - I}{h} - \mathcal{A}\right)f\right](w) + \lim_{h\searrow 0}\left(T_{s-h} - T_s\right)\mathcal{A}f(w) \\ &= \lim_{h\searrow 0}\mathbb{E}\left[\left(\frac{T_h - I}{h} - \mathcal{A}\right)f(we^{-s+h} + \sigma(s-h)Z)\right] \\ &\quad + \lim_{h\searrow 0}\mathbb{E}\left[\mathcal{A}f(we^{-s+h} + \sigma(s-h)Z) - \mathcal{A}f(we^{-s} + \sigma(s)Z)\right] \\ &\overset{(4.7)}{=} 0\end{aligned}$$

again, by dominated convergence. It can be applied because of (4.7) and the observation that for any $z \in D[0,1]$ and $h \in [0,1]$:

$$\begin{aligned}&\left|\mathcal{A}f(we^{-s+h} + \sigma(s-h)z) - \mathcal{A}f(we^{-s} + \sigma(s)z)\right| \\ =&\Big|-Df(we^{-s+h} + \sigma(s-h)z)[we^{-s+h} + \sigma(s-h)z] \\ &+ \mathbb{E}D^2 f(we^{-s+h} + \sigma(s-h)Z)[Z^{(2)}] \\ &-Df(we^{-s} + \sigma(s)z)[we^{-s} + \sigma(s)z] + \mathbb{E}D^2 f(we^{-s} + \sigma(s)z)[Z^{(2)}]\Big| \\ \leq&\|f\|_M\left(1 + \|we^{-s+h} + \sigma(s-h)z\|^2\right)\|we^{-s+h} + \sigma(s-h)z\| \\ &+ \|f\|_M\left(1 + \|we^{-s+h} + \sigma(s-h)z\|\right)\mathbb{E}\|Z\|^2 \\ &+ \|f\|_M\left(1 + \|we^{-s} + \sigma(s)z\|\right)\|we^{-s} + \sigma(s)Z\| + \|f\|_M\left(1 + \|we^{-s} + \sigma(s)z\|\right)\mathbb{E}\|Z\|^2 \\ \leq&\|f\|_M\left(1 + 2\|w\|^2 e^{-2s+2} + 2\sigma^2(s-1)\|z\|^2\right)\left(\|w\|e^{-s+1} + \sigma(s-1)\|z\|\right) \\ &+ \|f\|_M\left(1 + \|we^{-s+1} + \sigma(s-1)z\|\right)\mathbb{E}\|Z\|^2 \\ &+ \|f\|_M\left(1 + \|we^{-s} + \sigma(s)z\|\right)\|we^{-s} + \sigma(s)z\| + \|f\|_M\left(1 + \|we^{-s} + \sigma(s)z\|\right)\mathbb{E}\|Z\|^2\end{aligned}$$

and so for any $h \in [0,1]$, $\left|\mathcal{A}f(we^{-s+h} + \sigma(s-h)Z) - \mathcal{A}f(we^{-s} + \sigma(s)Z)\right|$ is bounded by a random variable with finite expectation.

Thus, for all $w \in D[0,1]$ and $s > 0$:

$$\frac{d}{ds}T_s f(w) = T_s \mathcal{A} f(w)$$

and so, by the Fundamental Theorem of Calculus:

$$T_r f(w) - f(w) = \int_0^r T_s \mathcal{A} f(w) ds. \tag{4.8}$$

By Remark 4.2, we can apply (4.8) to $f = \int_0^t T_u g du$ to obtain:

$$T_r \int_0^t T_u g(w) du - \int_0^t T_u g(w) du = \int_0^r T_s \mathcal{A} \left(\int_0^t T_u g(w) du \right) ds.$$

Now, we take $t \to \infty$. Let Z' be an independent copy of Z. We apply dominated convergence, which is allowed because of (4.2) and the following bound for $\varphi_t(w) = \int_0^t T_u g(w) du$:

$$\begin{aligned}
&|\mathcal{A}\varphi_t(w)| \\
\leq &\int_0^t \mathbb{E}_Z \left| e^{-u} Dg(we^{-u} + \sigma(u)Z)[w] \right| du \\
&+ \int_0^t \mathbb{E}_Z \left\{ \mathbb{E}_{Z'} \left| e^{-2u} D^2 g(we^{-u} + \sigma(u)Z) \left[(Z')^{(2)} \right] \right| \right\} du \\
\leq &\int_0^\infty \mathbb{E}_Z \left| e^{-u} Dg(we^{-u} + \sigma(u)Z)[w] \right| du \\
&+ \int_0^\infty \mathbb{E}_Z \left\{ \mathbb{E}_{Z'} \left| e^{-2u} D^2 g(we^{-u} + \sigma(u)Z) \left[(Z')^{(2)} \right] \right| \right\} du \\
\leq &\|g\|_M \int_0^\infty e^{-u} \left(1 + \mathbb{E}_Z \|we^{-u} + \sigma(u)Z\|^2 \right) \|w\| du \\
&+ \|g\|_M \int_0^\infty e^{-2u} \left(1 + \mathbb{E}_Z \|we^{-u} + \sigma(u)Z\| \right) \mathbb{E}_{Z'} \|Z'\|^2 du \\
\leq &\|g\|_M \int_0^\infty \left(e^{-u} + 2\|w\|^2 e^{-3u} + 2\mathbb{E}_Z \|Z\| (e^{-u} - e^{-3u}) \right) \|w\| du \\
&+ \|g\|_M \int_0^\infty \left(e^{-2u} + \|w\| e^{-3u} + \sigma(u) e^{-2u} \right) \mathbb{E}_Z \|Z\|^2 du \\
\leq &\left(1 + \frac{4}{3} \mathbb{E}_Z \|Z\|^2 \right) \|g\|_M (1 + \|w\|^2) \|w\| + \left(\frac{1}{2} + \frac{\mathbb{E}_Z \|Z\|}{3} \right) \|g\|_M (1 + \|w\|) \mathbb{E}_Z \|Z\|,
\end{aligned}$$

where the second inequality follows again by dominated convergence applied because of (4.2) in order to exchange integration and differentiation in a way similar to (4.3). Then, we obtain:

$$\begin{aligned}
T_r \int_0^\infty T_u g(w) du - \int_0^\infty T_u g(w) du &= \int_0^r T_s \lim_{t\to\infty} \mathcal{A} \left(\int_0^t T_u g(w) du \right) ds \\
&\overset{(4.4)}{=} - \int_0^r T_s g(w) ds.
\end{aligned}$$

Now, by Lemma 4.1, we can divide both sides by r and take $r \to 0$ to obtain:

$$\begin{aligned}
\mathcal{A} \left(\int_0^\infty T_u g(w) du \right) &= - \lim_{r\to 0} \frac{1}{r} \int_0^r T_s g(w) ds = - \lim_{r\to 0} \left[\frac{1}{r} \int_0^r \mathbb{E} g(we^{-s} + \sigma(s)Z) ds \right] \\
&\overset{(4.6)}{=} -g(w),
\end{aligned}$$

which finishes the proof. □

Remark 4.5. In [6, Proposition 15] the authors prove that the semigroup of an $\mathbb{R}^d$-valued Itô diffusion with Lipschitz drift and diffusion coefficients is strongly continuous on the space $L' = \{x \mapsto (1+\|x\|^2)f(x) : f \in C_0(\mathbb{R}^d)\}$, equipped with the norm $\|f\|_{L'} = \sup_{x\in\mathbb{R}^d} \|f(x)\|_2/(1+\|x\|_2)$, where $C_0(\mathbb{R}^d)$ is the set of all continuous functions vanishing at infinity and $\|\cdot\|_2$ is the l^2 norm on $\mathbb{R}^d$. It might seem natural to try to adapt their argument to the infinite-dimensional setting and consider the space $L'' = \{w \mapsto (1+\|w\|^4)f(w) : f \in C_0(D,\mathbb{R})\}$, equipped with the norm $\|f\|_{L''} = \sup_{w\in D} |f(w)|/(1+\|w\|^4)$. Since $M \subset L'' \subset L$, the semigroup 2.1 being strongly continuous on L'' would readily imply Proposition 4.4.

However, there is no easy extension of the argument used in the proof of [6, Proposition 15] to the infinite dimensional setting. The reason is that the Riesz-Markov theorem for space L' [4, Theorem 2.4] invoked in the proof, requires a closed unit ball in the domain of the functions in L' to be compact. In other words, it requires the domain of the functions in L' to be a finite-dimensional space. Since D is infinite-dimensional, [4, Theorem 2.4] cannot be easily adapted to our setting and so the proof of [6, Proposition 15] cannot be easily adapted either.

References

[1] A.D. Barbour, *Stein's method and Poisson process convergence*, Journal of Applied Probability **25** (1988), 175–184.

[2] A.D. Barbour, *Stein's Method for Diffusion Approximations*, Probability Theory and Related Fields **84** (1990), 297–322.

[3] Giuseppe Daprato and Alessandra Lunardi, *On the Ornstein-Uhlenbeck operator in spaces of continuous functions*, Journal of Functional Analysis **131** (1995), no. 1, 94–114.

[4] P. Doersek and J. Teichmann, *A Semigroup Point of View On Splitting Schemes For Stochastic (Partial) Differential Equations*, arXiv:1011.2651, 2010.

[5] S.N. Ethier and T.G. Kurtz, *Markov processes: characterization and convergence*, Wiley, New York, 1986.

[6] J. Gorham, A.B. Duncan, S.J. Vollmer, and L. Mackey, *Measuring sample quality with diffusions*, arXiv:1611.06972, 2016.

[7] S. Holmes and G. Reinert, *Stein's method for bootstrap*, Lecture Notes-Monograph Series, vol. 46, Institute of Mathematical Statistics, 2004.

[8] Luigi Manca, *Kolmogorov operators in spaces of continuous functions and equations for measures*, Ph.D. thesis, Scuola Normale Superiore di Pisa, 2008.

Acknowledgments. We would like to thank Professor A.D. Barbour for an interesting discussion related to this work.

Statement of Authorship for joint/multi-authored papers for PGR thesis
To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	Note on A. Barbour's paper on Stein's method for diffusion approximations
Publication Status	☒ Published □ Accepted for Publication □ Submitted for Publication □ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Kasprzak, Mikołaj J.; Duncan, Andrew B.; Vollmer, Sebastian J. Note on A. Barbour's paper on Stein's method for diffusion approximations. Electron. Commun. Probab. 22 (2017), paper no. 23, 8 pp. doi:10.1214/17-ECP54.

Student Confirmation

Student Name:	Mikolaj Kasprzak		
Contribution to the Paper	Made the idea of the co-authors for a counterexample to strong continuity concrete. Constructed the entire proof of the validity of the main results of A. Barbour's paper holding despite the mistake and applied co-authors' ideas for minor corrections therein.		
Signature	Mikołaj Kasprzak.	Date	1 Oct 2018

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title:	Prof. GESINE REINERT		
Supervisor comments	Mikolaj's contribution to the paper are, to the best of my knowledge, well described.		
Signature	Gesine Reinert	Date	1 Oct 2018

This completed form should be included in the thesis, at the end of the relevant chapter.

Diffusion approximations via Stein's method and time changes

Mikołaj J. Kasprzak,

Mikołaj Kasprzak
Department of Statistics
University of Oxford
24-29 St Giles'
Oxford, OX1 3LB
United Kingdom
e-mail: kasprzak@stats.ox.ac.uk

Abstract: We extend the ideas of [Bar90] and use Stein's method to obtain a bound on the distance between a scaled time-changed random walk and a time-changed Brownian Motion. We then apply this result to bound the distance between a time-changed compensated scaled Poisson process and a time-changed Brownian Motion. Finally, we study the rate of convergence of the law of a rescaled M/M/1 queue to that of a time-changed Brownian Motion. Our approach may be extended to a wider class of continuous time Markov chains whose jump rates do not depend on the current state yet only on the jump sizes.

MSC 2010 subject classifications: Primary 60B10, 60F17; secondary 60J70, 60J65, 60E05, 60E15.
Keywords and phrases: Stein's method, functional convergence, time-changed Brownian Motion, M/M/1 queue.

1. Introduction

In the seminal paper [Bar90], Barbour observed that the celebrated Stein's method, first introduced in [Ste72] as a tool for proving the Central Limit Theorem, may also be used in the setup of the *Functional* Central Limit Theorem. The method provides a framework for proving distributional convergence results, together with bounds on the rate of this convergence. Before [Bar90] it had only been applied to finite-dimensional probability distributions. Barbour, in his work, looked at weak convergence to the Wiener measure. He considered Donsker's theorem which says that for a sequence of i.i.d. random variables $(X_n)_{n=1}^{\infty}$ with mean zero and unit variance, the random process

$$t \mapsto \sum_{i=1}^{\lfloor nt \rfloor} X_i$$

defined for $t \in [0,1]$ converges in distribution (as $n \to \infty$) to the standard Brownian Motion with respect to the Skorokhod topology. In this paper we shall extend Barbour's approach to weak convergence to a time-changed Brownian

Motion of birth and death processes whose jump rates do not depend on the current state. The particular example we consider is that of a rescaled M/M/1 queue.

1.1. Motivation

Functional limit results are central in many applied fields. Continuous processes arising as scaling limits of discrete ones are often easier to study and more robust to local changes than the processes they approximate. This is why they are often chosen as models for real-life phenomena, even those which are discrete in nature. Obtaining bounds on the rate of functional convergence is of great importance in determining the quality of this choice. Our motivation in this paper comes from a desire to fill in a gap in the literature but has also direct relevance to applications.

The M/M/1 queue, which we study in this paper, is used in applied fields whenever arrivals and departures from an operational system with one server need to be modelled. An account of its properties and applications may be found, for instance, in [Gau12]. Some of the example uses of the model include organising the staffing and work assignments in a call centre or hospital emergency ward planning. Its diffusion approximation might provide some intuition as to how the parameters of the queue impact some performance measures and thus help in optimisation exercises.

In [BDM18], the authors derive a bound on a distance between a scaled M/M/1 queue and a Brownian Motion. The bound is derived with respect to a distance constructed in a very technical and functional-analysis-heavy way. The distance and the bound are, moreover, not convergence-determining. In other words, even though the bound in [BDM18] converges to zero as $n \to \infty$, this does not imply weak convergence of the rescaled M/M/1 queue to the Wiener measure. In Remark 3.9 we explain in detail the setup and results of [BDM18].

Motivated by the wide applicability of the model, our aim is to derive a bound with respect to an easily interpretable notion of distance and, crucially, one that is strong enough to metrize weak convergence.

1.2. Contribution of the paper

The main achievements of the paper are the following:

1. An extension of the result of [Bar90] to time-changed scaled random walks. Specifically, we consider a sequence of i.i.d. random variables $X_1, X_2, \cdots$ with mean zero and unit variance and the process:

$$\mathbf{Y}_n(t) = n^{-1/2} \sum_{i=1}^{\lfloor nS(t) \rfloor} X_i, \quad t \in [0,1],$$

   for some deterministic time change S. We also consider a time-changed Brownian Motion $\mathbf{Z} = \mathbf{B} \circ S$, for a standard Brownian Motion $\mathbf{B}$. In

Theorem 3.1 we bound the following quantity:

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{Z})|$$

for every bounded, twice differentiable real-valued function g on the Skorokhod space, whose first two derivatives are uniformly bounded and the second derivative is Lipschitz.

2. A quantitative version of the functional limit theorem for (time-changed) compensated Poisson processes. We consider a sequence of deterministic time changes $\left(S^{(n)}\right)_{n\geq 1}$ converging to a limiting time change S. We then look at a rate 1 Poisson process $\mathbf{P}$ and bound the distance between its compensated, scaled, time-changed version:

$$\frac{\mathbf{P}(nS^{(n)}(t)) - nS^{(n)}(t)}{\sqrt{n}}, \quad t \in [0,1],$$

and the time-changed Brownian Motion $\mathbf{B}\circ S$. We do so for the same class of test functions as in Theorem 3.1. This result is presented in Theorem 3.4.

3. A quantitative version of a functional limit theorem for the M/M/1 queue. We denote by $\mathbf{L}(t)$ the length of an M/M/1 queue with arrival and departure rates $n\lambda$ and $n\mu$, at time t, and by $\overline{\mathbf{L}}$ its *fluid limit*:

$$\overline{\mathbf{L}}(t) = (x + (\lambda - \mu)t)^{+}.$$

We consider

$$\mathbf{Y}_n(t) = \sqrt{n}\left(\frac{\mathbf{L}(nt)}{n} - \overline{\mathbf{L}}(t)\right), \quad t \in [0,1].$$

and, in Theorem 3.7, bound its distance from a time-changed Brownian Motion, under the assumption that the starting point of $\overline{\mathbf{L}}$ is greater than $n(\mu - \lambda)$. We do so for the same test functions as those considered in the previous points. Our approach to proving this result may be extended to a wider-class of continuous-time Markov chains whose jump rates do not depend on the state the process is in, as described in Remark 5.1.

4. An extension of [BJ09, Propoision 3.1], which shows that the class of functions g which are bounded, twice differentiable with their derivatives bounded and whose second derivative is Lipschitz is rich enough to metrize weak convergence. Hence, weak convergence results follow from the bounds obtained in Theorems 3.1, 3.4 and 3.7. This is obtained in Propositions 2.1 and 2.2.

1.3. Stein's method

The aim of Stein's method is to find a bound on the quantity $|\mathbb{E}_{\nu_n}h - \mathbb{E}_{\mu}h|$, where μ is the target (known) distribution, ν_n is the approximating law and h

is chosen from a suitable class of real-valued test functions $\mathcal{H}$. The idea is to find an operator $\mathcal{A}$ acting on a class of real-valued functions such that

$$(\forall f \in \text{Domain}(\mathcal{A}) \quad \mathbb{E}_\nu \mathcal{A}f = 0) \quad \Longleftrightarrow \quad \nu = \mu,$$

where μ is the target distribution. In the next step, for a given function $h \in \mathcal{H}$, a solution $f = f_h$ to the following Stein equation:

$$\mathcal{A}f = h - \mathbb{E}_\mu h$$

is sought and its properties studied. Finally, using various mathematical tools (among which the most popular are Taylor's expansions in the continuous case, Malliavin calculus, as described in [NP12], and coupling methods), a bound is sought for the quantity $|\mathbb{E}_{\nu_n}\mathcal{A}f_h|$.

An accessible account of the method can be found, for example, in the surveys [LRS17] and [Ros11] as well as the books [BHJ92] and [CGS11], which treat the cases of Poisson and normal approximation, respectively, in detail. The reference [Swa16] is a database of information and publications connected to Stein's method.

Approximations by infinite-dimensional laws have not been covered in the Stein's method literature very widely, with the notable exceptions of [Bar90, BJ09, CD13] and, more recently, [Kas18, Kas17, BDM18]. In this paper, we will focus on the ideas taken from [Bar90] and [BDM18].

1.4. Structure of the paper and notation

In Section 2 we introduce the space of test functions we will find the bounds for. In Section 3 we present our main results. Theorem 3.1 shows how the approach in [Bar90] can be extended to the approximation of a scaled, time-changed random walk by a time-changed Brownian Motion. In Theorem 3.4 we apply Theorem 3.1 to look at the distance between a time-changed Poisson Process and a time-changed Browian Motion. Theorem 3.7 treats the M/M/1 queue. Section 4 sets up Stein's method for proving the results and Section 5 provides the actual proofs.

In what follows, $\|\cdot\|$ will always denote the sup norm and $D = D[0,1] = D([0,1],\mathbb{R})$ will be the Skorokhod space of càdlàg real-valued functions on $[0,1]$. By $D^k f$ we mean the k-th Fréchet derivative of f and the norm of a k-linear form B on L is defined to be $\|B\| = \sup_{\{h:\|h\|=1\}} |B[h,...,h]|$.

2. Space M^0

Let M^0 be the class of functionals $g : D[0,1] \to \mathbb{R}$ such that:

$$\|g\|_{M^0} := \sup_{w\in D} |g(w)| + \sup_{w\in D} \|Dg(w)\| + \sup_{w\in D} \|D^2 g(w)\| + \sup_{w,h\in D} \frac{\|D^2 g(w+h) - D^2 g(w)\|}{\|h\|} < \infty.$$

This is Proposition 3.1 of [BJ09]:

Proposition 2.1. *Suppose that, for each $n \geq 1$, the random element $\mathbf{Y}_n$ of $D[0,1]$ is piecewise constant with intervals of constancy of length at least r_n. Let $(\mathbf{Z}_n)_{n\geq 1}$ be random elements of $D[0,1]$ converging weakly in $D[0,1]$, with respect to the Skorokhod topology, to a random element $\mathbf{Z} \in C([0,1],\mathbb{R})$. If:*

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{Z}_n)| \leq C\tau_n \|g\|_{M^0} \tag{2.1}$$

for each $g \in M^0$ and if $\tau_n \log^2(1/r_n) \xrightarrow{n\to\infty} 0$, then $\mathbf{Y}_n$ converges in distribution to $\mathbf{Z}$ with respect to both the Skorokhod and the uniform topologies.

A similar result holds when $\mathbf{Y}_n$ is a continuous-time Markov chain:

Proposition 2.2. *Suppose that, for each $n \geq 1$, the random element $\mathbf{Y}_n$ of $D[0,1]$ is a contiuous-time Markov chain with mean holding time $\frac{1}{\lambda_n} \to 0$. Let $(\mathbf{Z}_n)_{n\geq 1}$ be random elements of $D[0,1]$ converging weakly in $D[0,1]$, with respect to the Skorokhod topology, to a random element $\mathbf{Z} \in C([0,1],\mathbb{R})$. Suppose further that:*

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{Z}_n)| \leq C\tau_n \|g\|_{M^0} \tag{2.2}$$

for each $g \in M^0$ and that $\tau_n \log^2(\lambda_n) \xrightarrow{n\to\infty} 0$. Then $\mathbf{Y}_n$ converges in distribution to $\mathbf{Z}$ with respect to both the Skorokhod and the uniform topologies.

We prove Proposition 2.2 in Appendix A.

3. Main results

Theorem 3.1 below is an extension of [Bar90, Theorem 1] to the case of a time-changed scaled random walk:

Theorem 3.1. *Let $X_1, X_2, \ldots$ be i.i.d. with mean 0, variance 1 and finite third moment. Let $S : [0,1] \to [0,\infty)$ be a strictly increasing, continuous function with $S(0) = 0$. Define:*

$$\mathbf{Y}_n(t) = n^{-1/2} \sum_{i=1}^{\lfloor nS(t) \rfloor} X_i, \quad t \in [0,1]$$

and let $(\mathbf{Z}(t), t \in [0,1]) = (\mathbf{B}(S(t)), t \in [0,1])$, where $\mathbf{B}$ is a standard Brownian Motion. Suppose that $g \in M^0$, as defined in Section 2. Then:

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{Z})| \leq \frac{\|g\|_{M^0}}{n^{1/2}} \left(\frac{30}{\sqrt{\pi \log 2}} \sqrt{\log(2S(1)n)} + S(1)\mathbb{E}|X_1|^3 \right).$$

Remark 3.2 (Relevance of terms in the bound)**.** *In the proof of Theorem 3.1, the distribution of $\mathbf{Y}_n$ is first compared to the distribution of a piecewise constant Gaussian process with the same covariance structure as that of $\mathbf{Y}_n$. This comparison gives rise to the second term in the bound: $\frac{\|g\|_{M^0}}{n^{1/2}} S(1)\mathbb{E}|X_1|^3$, which is a Berry-Esseen-type term involving the third absolute moment of the summands.*

The piecewise constant Gaussian process is then compared to the time-changed Brownian Motion and hence the term of order $\frac{\sqrt{\log n}}{\sqrt{n}}$ *arises. It corresponds to the transition from a càdlàg process to a continuous one and is calculated using the Brownian modulus of continuity.*

Remark 3.3. *In Theorem 3.1 we do not claim that our bounds are sharp. Our bound in Theorem 3.1 is of the same order as the one obtained in the original case in [Bar90]. This result can also be extended in a straightforward way to instances in which the time change is random and independent of the step sizes of the random walk. We can obtain this by conditioning on the time change.*

Theorem 3.4 below treats a time-changed Poisson process and can also be extended to random time changes, independent of the Poisson process of interest, by conditioning.

Theorem 3.4. *Suppose that* $\mathbf{P}$ *is a Poisson process with rate* 1 *and* $S^{(n)}: [0,1] \to [0,\infty)$ *is a sequence of increasing continuous functions, such that* $S^{(n)}(0) = 0$*. Let* $S : [0,1] \to [0,\infty)$ *be also increasing and continuous. Let* $\mathbf{Z}(t) = \mathbf{B}(S(t)), t \in [0,1]$ *where* $\mathbf{B}$ *is a standard Brownian Motion and*

$$\tilde{\mathbf{Y}}_n(t) = \frac{\mathbf{P}\left(nS^{(n)}(t)\right) - nS^{(n)}(t)}{\sqrt{n}}, \qquad t \in [0,1].$$

Then, for all $g \in M^0$ *and* $n \geq 3$*,*

$$\begin{aligned}
&|\mathbb{E}g(\tilde{\mathbf{Y}}_n) - \mathbb{E}g(\mathbf{Z})| \\
\leq &\frac{\|g\|_{M^0}}{n^{1/2}}\left(\frac{30}{\sqrt{\pi \log 2}}\sqrt{\log(2S(1)n)} + 2S^{(n)}(1) + 1 + \frac{-1+2\log n}{\log\log n}\right) \\
&+ 2\|g\|_{M^0}\sqrt{\|S - S^{(n)}\|}.
\end{aligned}$$

Remark 3.5 (Relevance of terms in the bound)**.** *In the proof of Theorem 3.4, the distribution of* $\tilde{\mathbf{Y}}_n$ *is first compared to the distribution of a scaled time-changed Poisson random walk and this is where the term* $\frac{\|g\|_{M^0}}{n^{1/2}} \cdot \frac{-1+2\log n}{\log\log n}$ *comes from. The last term in the bound arises because of the transition between the pre-limiting sequence of time changes and the limiting time change. The remaining terms are incorporated from Theorem 3.1 applied to the Poisson random walk.*

Remark 3.6. *The bound in Theorem 3.4 goes to* 0 *as long as the time changes* $S^{(n)} \to S$ *uniformly. By Proposition 2.2 it then follows that* $\mathbf{Y}_n$ *converges to* $\mathbf{Z}$ *in distribution with respect to both Skorokhod and uniform topologies.*

Theorem 3.7 below establishes a bound on the rate of convergence in a functional limit theorem for the M/M/1 queue. In what follows we will use the setup and notation of [Rob03, Chapter 5.7]. For two independent rate 1 Poisson processes $\mathbf{P}_1$ and $\mathbf{P}_{-1}$, the M/M/1 queue with jump rates λ and μ is the solution to the following equation:

$$\mathbf{L}(t) = \mathbf{L}(0) + \mathbf{P}_1(\lambda t) - \int_0^t \mathbb{1}_{\{\mathbf{L}(s-)>0\}}\mathbf{P}_{-1}(\mu ds). \tag{3.1}$$

Let $(x_n)_{n\geq 1}$ and x be such that $\frac{x_n}{n}\xrightarrow{n\to\infty} x\in\mathbb{R}_+$. Consider the renormalised proces

$$\overline{\mathbf{L}_n}(t)=\frac{\mathbf{L}(nt)}{n},\quad t\in[0,1],\tag{3.2}$$

where $(\mathbf{L}(t),t\in[0,1])$ solves (3.1) with initial condition x_n. The first part of [Rob03, Proposition 5.16] states that $\overline{\mathbf{L}_n}$ of (3.2) converges with respect to uniform topology to the function $t\mapsto\overline{\mathbf{L}}(t):=(x+(\lambda-\mu)t)^+$. The second part of the Proposition states that, if $\sqrt{n}(\frac{x_n}{n}-x)\xrightarrow{n\to\infty}0$ then the process

$$\mathbf{Y}_n(t)=\sqrt{n}\left(\overline{\mathbf{L}_n}(t)-\overline{\mathbf{L}}(t)\right),\quad t\in[0,1]$$

converges in distribution to $(\mathbf{B}\left((\lambda+\mu)t\right),t\in[0,1])$, where $\mathbf{B}$ is a standard Brownian Motion.

We prove the following theorem

Theorem 3.7 (M/M/1 queue)**.** *Let $\mathbf{P}_1$ and $\mathbf{P}_{-1}$ be two independent rate 1 Poisson processes. Let $\lambda,\mu,x\geq 0$ and $n\geq 3$. Suppose that $x\geq\mu-\lambda$ and that $(\mathbf{L}(t),t\in[0,1])$ solves equation (3.1) with initial condition nx. Consider processes given by*

$$\overline{\mathbf{L}_n}(t)=\frac{\mathbf{L}(nt)}{n},\quad \overline{\mathbf{L}}(t)=(x+(\lambda-\mu)t)^+,\quad t\in[0,1].$$

Let $\mathbf{B}$ be a standard Brownian Motion and let $\mathbf{Z}(t)=\mathbf{B}\left((\lambda+\mu)t\right)$ for $t\in[0,1]$. Consider

$$\mathbf{Y}_n(t)=\sqrt{n}\left(\overline{\mathbf{L}_n}(t)-\overline{\mathbf{L}}(t)\right),\quad t\in[0,1].$$

Then, for any $g\in M^0$ and $n\geq 3$ we have

$$\begin{aligned}&|\mathbb{E}g(\mathbf{Y}_n)-\mathbb{E}g(\mathbf{Z})|\\ \leq&\frac{\|g\|_{M^0}}{n^{1/2}}\left(\frac{30\left(\sqrt{\log(2\lambda n)}+\sqrt{\log(2\mu n)}\right)}{\sqrt{\pi\log 2}}+2(\lambda+\mu)+2+\frac{2\log 2+4\log n}{\log\log(n+2)}\right)\\ &+c\|g\|_{M^0}e^{-n},\end{aligned}$$

for some constant c independent of n and g.

Remark 3.8 (Relevance of terms in the bound)**.** *In the proof of Theorem 3.7, we consider two cases. The first one is that the process $\overline{\mathbf{L}_n}$ hits 0 (or returns to 0 if started there) before time 1. It can be proved that, as long as $x\geq\mu-\lambda$, the probability of this happening is small. In this case, the bound takes the form $c\|g\|_{M^0}e^{-n}$ and hence this term appears in our general bound. In the opposite case, the process $\mathbf{Y}_n$ of Theorem 3.7 is a difference of two scaled compensated Poisson processes and an application of Theorem 3.4 gives rise to the remaining terms in the bound.*

Remark 3.9. *The recent paper [BDM18] also establishes a bound on the rate of convergence of $\mathbf{Y}_n$ of Theorem 3.7 to $\mathbf{Z}$ in the case of $x_n=nx$ and $x\geq\mu-\lambda$.*

However, the distance with respect to which the authors obtain their bound is different from the distance we consider. In order to understand the definition of the distance looked at in [BDM18] we need to introduce some notation.

First, for $\eta \in (0,1]$ and $p \geq 1$, we denote the corresponding fractional Sobolev space by $W_{\eta,p}$, which is defined as the closure of C^1 functions with respect to the norm:

$$\|f\|_{\eta,p}^p = \int_0^1 |f(t)|^p dt + \int\int_{[0,1]^2} \frac{|f(t)-f(s)|^p}{|t-s|^{1+p\eta}} dt\, ds.$$

Next, we introduce the Besov-Liouville spaces. For $f \in L^p([0,1],dt)$ (denoted by L^1 for short), the left fractional integral of f is defined by:

$$(I_{0+}^{\alpha})(x) = \frac{1}{\Gamma(\alpha)} \int_0^x f(t)(x-t)^{\alpha-1} dt, \quad x \geq 0,$$

where $\alpha > 0$ and $I_{0+}^0 = Id$. The Besov-Liouville space $\mathcal{I}_{\alpha,p}^+$ is defined by $I_{0+}^{\alpha}(L^p)$ and equipped with the norm

$$\|I_{0+}^{\alpha} f\|_{\mathcal{I}_{\alpha,p}^+} = \|f\|_{L^p}.$$

Moreover, we call a function $F : W_{\eta,p} \to \mathbb{R}$ cylindrical if it is of the form

$$F = f(\delta_B h_1, \cdots, \delta_B h_k),$$

where f belongs to the Schwartz space on $\mathbb{R}^k$, $h_1, \cdots, h_k$ belong to $\mathcal{I}_{1,2}^+$ and $\delta_B h$ is the Ito integral of h given by

$$\delta_B h = \int_0^1 \dot{h}(s) dB(s).$$

Now, for such a function F, we let ∇F be an element of $L^2(W_{\eta,p}, \mathcal{I}_{1,2}^+)$ defined by

$$\nabla F = \sum_{j=1}^k \partial_j f(\delta_B h_1, \cdots, \delta_B h_k) h_j$$

and $\nabla^{(2)}$ be the element of $L^2(W_{\eta,p}, \mathcal{I}_{1,2}^+ \otimes \mathcal{I}_{1,2}^+)$, given by

$$\nabla^{(2)} F = \sum_{j,l=1}^k \partial_{jl}^{(2)} f(\delta_B h_1, \cdots, \delta_B h_k) h_j \otimes h_l.$$

For such an F, we also consider the following norm:

$$\|F\|_{2,2}^2 = \|F\|_{L^2}^2 + \mathbb{E}\left[\|\nabla F\|_{\mathcal{I}_{1,2}^+}^2\right] + \mathbb{E}\left[\|\nabla^{(2)} F\|_{\mathcal{I}_{1,2}^+ \otimes \mathcal{I}_{1,2}^+}^2\right],$$

where

$$\|\nabla F\|_{\mathcal{I}_{1,2}^+}^2 = \int_0^1 \left(\sum_{j=1}^k \partial_j f(\delta_B h_1, \cdots, \delta_B h_k) \dot{h}_j(s)\right)^2 ds$$

and

$$\|\nabla^{(2)}F\|^2_{(\mathcal{I}^+_{1,2})^{\otimes 2}} = \int_0^1\int_0^1\left(\sum_{j,l=1}^k \partial^{(2)}_{jl} f(\delta_B h_1,\cdots,\delta_B h_k)\dot{h}_j(s)\dot{h}_k(s)\right)^2 ds\, dr.$$

We let $\mathbb{D}_{2,2}$ *be the completion of the set of cylindrical functions with respect to the norm* $\|\cdot\|_{2,2}$.

Furthermore, we define the class of functions $\Sigma\eta, p$ *as the collection of* 1-*Lipschitz functions* $F: W_{\eta,p}\to\mathbb{R}$ *which belong to* $\mathbb{D}_{2,2}$ *and satisfy*

$$\left|\left\langle \nabla^{(2)}F(x) - \nabla^{(2)}F(x+g), h\otimes k\right\rangle^+_{\mathcal{I}_{1,2}}\right| \le \|g\|_{W_{\eta,p}}\|h\|_{L^2}\|k\|_{L^2},$$

for any $x\in W_{\eta,p}$, $g\in\mathcal{I}^+_{1,2}$, $h,k\in L^2$.

Finally, we define the following space

$\mathscr{L}_{\eta,p} = \{$*bounded 1-Lipschitz functions* $D\to\mathbb{R}$ *whose restriction to* W_{η_p} *belongs to* $\Sigma_{\eta,p}\}$.

The authors of [BDM18] establish a bound on the following quantity

$$\sup_{F\in\mathscr{L}_{\eta,p}}\left|\mathbb{E}F(\mathbf{Y}_n) - \mathbb{E}F(\mathbf{Z})\right|. \tag{3.3}$$

The bound they obtained is of exactly the same order as our bound in Theorem 3.7 and the authors in [BDM18] do not provide any numerical values for the constants in their bound. However, the class of functions considered in [BDM18] does not metrize weak convergence; i.e. the bound obtained therein does not imply weak convergence of the law of $\mathbf{Y}_n$ *to that of* $\mathbf{Z}$. *Remark 3.11 below shows that our bound does.*

Remark 3.10. *The proof of the bound on the distance between the rescaled M/M/1 queue and the Wiener process presented in [BDM18] is based on Malliavin calculus on the Poisson space. This is one of the reasons why a complicated class of test functions, described in Remark 3.9, is considered therein.*

On the other hand, our proof of Theorem 3.7 uses Theorem 3.4 and an elementary Lindeberg-type trick which allows one to bound distances between sums of stochastic processes in case we know how to bound the distance between the corresponding summands. Indeed, the trick is used since an M/M/1 queue length conditioned to stay away from zero may be described as a difference of two Poisson processes looking at the arrivals and the departures, respectively. The limiting time-changed Brownian Motion may also easily be expressed as a difference of two time-changed independent Brownian Motions.

Remark 3.11. *By Proposition 2.2, Theorem 3.7 implies that* $\mathbf{Y}_n$ *converges in distribution to* $\mathbf{Z}$ *with respect to the Skorokhod and uniform topologies. Indeed, the mean holding time in the continuous-time Markov chain* $\mathbf{Y}_n$ *is of order* n^{-1} *and our bound is of order* $\frac{\log n}{n^{1/2}\log\log n}$. *As* $\frac{\log^3 n}{n^{1/2}\log\log n}\xrightarrow{n\to\infty} 0$, *Proposition 2.2 may be applied in this case.*

4. Setting up Stein's method

Let us first define:

$$\mathbf{A}_n(t) = n^{-1/2} \sum_{i=1}^{\lfloor nS(1) \rfloor} Z_i \mathbb{1}_{[i/n, S(1)]}(S(t)) = n^{-1/2} \sum_{i=1}^{\lfloor nS(1) \rfloor} Z_i \mathbb{1}_{[S^{-1}(i/n), 1]}(t), \tag{4.1}$$

for $Z_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$. In the proof of Theorem 3.1, we will apply Stein's method to find the distance between $\mathbf{A}_n$ and $\mathbf{Y}_n$.

4.1. The Stein equation

We first note that if $\mathscr{U}_1, \mathscr{U}_2, \ldots$ are i.i.d. Ornstein-Uhlenbeck processes with stationary law $\mathcal{N}(0,1)$, then defining:

$$\mathbf{W}_n(t,u) = n^{-1/2} \sum_{i=1}^{\lfloor nS(t) \rfloor} \mathscr{U}_i(u), \quad u \geq 0, t \in [0,1],$$

we obtain that the law of $\mathbf{A}_n$ is stationary for $(\mathbf{W}_n(\cdot,u))_{u\geq 0}$. Denote the generator of $(\mathbf{W}_n(\cdot,u))_{u\geq 0}$ by $\mathcal{A}_n$. By properties of stationary distributions, $\mathbb{E}_\mu \mathcal{A}_n f = 0$ for all $f \in \text{Domain}(\mathcal{A}_n)$ if and only if $\mu = \mathcal{L}(\mathbf{A}_n)$. Therefore, we can treat

$$\mathcal{A}_n f = g - \mathbb{E} g(\mathbf{A}_n) \tag{4.2}$$

as our Stein equation.

In the next subsection, for any g from a suitable class of functions, we will find an f satisfying equation (4.2). Then, in the sequel, we will find a bound on $|\mathbb{E}\mathcal{A}_n f(\mathbf{Y}_n)|$, which will readily give us a bound on $|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{A}_n)|$.

The following result is immediate by [Kas18, Proposition 4.1]:

Proposition 4.1. *The generator $\mathcal{A}_n$ of the process $(W_n(\cdot,u))_{u\geq 0}$ acts on any $f \in M^0$ in the following way:*

$$(\mathcal{A}_n f)(w) := -Df(w)[w] + \mathbb{E}D^2 f(w)\left[\mathbf{A}_n^{(2)}\right].$$

4.2. Solving the Stein equation

Proposition 4.2. *Suppose that $g \in M^0$ satisfies $\mathbb{E}g(\mathbf{A}_n) = 0$. Then the equation: $\mathcal{A}_n f = g$ is solved by:*

$$f = \phi_n(g) = -\int_0^\infty T_{n,u} g du, \tag{4.3}$$

where $(T_{n,u} f)(w) = \mathbb{E}\left[f(we^{-u} + \sqrt{1-e^{-2u}}\mathbf{A}_n)\right]$.

Furthermore, $\phi_n(g) \in M^0$ and the following inequalities hold:

$$\text{A)} \quad \|D\phi_n(g)(w)\| \leq \|g\|_{M^0},$$
$$\text{B)} \quad \|D^2\phi_n(g)(w)\| \leq \frac{\|g\|_{M^0}}{2},$$
$$\text{C)} \quad \|D^2\phi_n(g)(w+h) - D^2\phi(g)(w)\| \leq \frac{\|g\|_{M^0}}{3}\|h\|. \tag{4.4}$$

Proof. The fact that $\phi_n(g)$ of 4.3 solves the Stein equation $\mathcal{A}_n f = g$ follows by the argument used to prove [KDV17, Proposition 1] upon noting that we can readily substitute $\mathbf{A}_n$ is the place of $\mathbf{Z}$ therein.

Now, note that for ϕ_n of 4.3 and any $g \in M^0$, we get

$$\begin{aligned}&\phi_n(g)(w+h) - \phi_n(g)(w)\\ =& -\mathbb{E}\int_0^\infty \left[g\left((w+h)e^{-u} + \sqrt{1-e^{-2u}}\mathbf{A}_n\right) - g\left(we^{-u} + \sqrt{1-e^{-2u}}\mathbf{A}_n\right)\right] du\end{aligned}$$

and so dominated convergence (which can be applied because of [KDV17, (10)]) gives:

$$D^k\phi_n(g)(w) = -\mathbb{E}\int_0^\infty e^{-ku} D^k g\left(we^{-u} + \sqrt{1-e^{-2u}}\mathbf{A}_n\right) du, \quad k = 1, 2. \tag{4.5}$$

Using (4.5), we obtain that, for any $g \in M^0$ and $w, h \in D[0,1]$,

$$\text{A)} \quad \|D\phi_n(g)(w)\| \leq \int_0^\infty e^{-u}\mathbb{E}\left\|Dg\left(we^{-u} + \sqrt{1-e^{-2u}}\mathbf{A}_n\right)\right\| du \leq \|g\|_{M^0}$$

$$\text{B)} \quad \|D^2\phi_n(g)(w)\| \leq \int_0^\infty e^{-2u}\mathbb{E}\left\|D^2 g\left(we^{-u} + \sqrt{1-e^{-2u}}\mathbf{A}_n\right)\right\| du \leq \frac{\|g\|_{M^0}}{2}$$

$$\begin{aligned}\text{C)} \quad & \frac{\|D^2\phi_n(g)(w+h) - D^2\phi_n(g)(w)\|}{\|h\|}\\ \leq & \|h\|^{-1}\left\|\mathbb{E}\int_0^\infty e^{-2u}\left[D^2 g\left((w+h)e^{-u} + \sqrt{1-e^{-2u}}\mathbf{A}_n\right)\right.\right.\\ & \left.\left. - D^2 g\left(we^{-u} + \sqrt{1-e^{-2u}}\mathbf{A}_n\right)\right] du\right\|\\ \leq & \frac{\|g\|_{M^0}}{3}.\end{aligned}$$

□

Remark 4.3. *It is an easy consequence of Propositions 4.1 and 4.2 that for $g \in M^0$:*

$$\mathcal{A}_n\phi_n(g)(w) = -D\phi_n(g)(w)[w] + \mathbb{E}D^2\phi_n(g)(w)\left[\mathbf{A}_n^{(2)}\right].$$

5. Proofs of Theorems 3.1, 3.4 and 3.7

5.1. Proof of Theorem 3.1

5.1.1. Discretisation of Brownian Motion

Let $\mathbf{A}_n$ be as in (4.1). Now, note that we can first realise a standard Brownian Motion $\mathbf{B}$ and then set $\mathbf{A}_n(t)=\mathbf{B}\left(\frac{\lfloor nS(t)\rfloor}{n}\right)$ and $\mathbf{Z}=\mathbf{B}$ for $t\in[0,1]$ so that:

$$\sup_{t\in[0,1]}|\mathbf{A}_n(t)-\mathbf{Z}(t)|=\sup_{t\in[0,1]}\left|\mathbf{B}\left(\frac{\lfloor nS(t)\rfloor}{n}\right)-\mathbf{B}(S(t))\right|=\sup_{t\in[0,S(1)]}\left|\mathbf{B}(t)-\mathbf{B}\left(\frac{\lfloor nt\rfloor}{n}\right)\right|.$$

By [FN10, Lemma 3] we get:

$$\mathbb{E}\|\mathbf{A}_n-\mathbf{Z}\|\leq\mathbb{E}\left[\sup_{t,s\in[0,S(1)],|t-s|\leq\frac{1}{n}}|\mathbf{B}(t)-\mathbf{B}(s)|\right]\leq\frac{30}{\sqrt{\pi\log 2}}\frac{\sqrt{\log(2S(1)n)}}{\sqrt{n}} \tag{5.1}$$

and therefore we obtain, for any $g\in M^0$, as defined in Section 2,

$$\begin{aligned}|\mathbb{E}g(\mathbf{A}_n)-\mathbb{E}g(\mathbf{Z})|&\overset{\text{MVT}}{\leq}\mathbb{E}\left[\sup_{c\in[0,1]}\|Dg\left((1-c)\mathbf{Z}+c\mathbf{A}_n\right)\|\|\mathbf{Z}-\mathbf{A}_n\|\right]\\&\leq\|g\|_{M^0}\mathbb{E}\|\mathbf{Z}-\mathbf{A}_n\|\\&\leq\frac{30\|g\|_{M^0}}{\sqrt{\pi\log 2}}\frac{\sqrt{\log(2S(1)n)}}{\sqrt{n}}.\end{aligned} \tag{5.2}$$

5.1.2. Applying Stein's method

Let $g\in M^0$ and $g_n=g-\mathbb{E}[g(\mathbf{A}_n)]$. Let $f_n=\phi_n(g_n)$, as in (4.3). First, note that:

$$\mathbb{E}Df_n(\mathbf{Y}_n)[\mathbf{Y}_n]=n^{-1/2}\sum_{j=1}^{\lfloor nS(1)\rfloor}\mathbb{E}Df_n(\mathbf{Y}_n)[X_j\mathbb{1}_{[S^{-1}(j/n),1]}].$$

We now let $\mathbf{Y}_n^j=n^{-1/2}\sum_{k\neq j}X_k\mathbb{1}_{[S^{-1}(k/n),1]}=\mathbf{Y}_n-n^{-1/2}X_j\mathbb{1}_{[S^{-1}(j/n),1]}$ and observe that, by Taylor's theorem:

$$\begin{aligned}&\Big|n^{-1/2}\mathbb{E}X_jDf_n(\mathbf{Y}_n)\left[\mathbb{1}_{[S^{-1}(j/n),1]}\right]-\mathbb{E}\Big\{n^{-1/2}X_jDf_n(\mathbf{Y}_n^j)\left[\mathbb{1}_{[S^{-1}(j/n),1]}\right]\\&+n^{-1}(X_j)^2D^2f_n(\mathbf{Y}_n^j)\left[\left(\mathbb{1}_{[S^{-1}(j/n),1]}e_i\right)^{(2)}\right]\Big\}\Big|\\=&\Big|\mathbb{E}\Big[n^{-1/2}X_jDf_n\left(\mathbf{Y}_n^j+n^{-1/2}X_j\mathbb{1}_{[j/n,1]}\right)\left[\mathbb{1}_{[S^{-1}(j/n),1]}\right]\\&-n^{-1/2}X_jDf_n(\mathbf{Y}_n^j)\left[\mathbb{1}_{[S^{-1}(j/n),1]}\right]\\&-n^{-1}(X_j)^2D^2f_n(\mathbf{Y}_n^j)\left[\left(\mathbb{1}_{[S^{-1}(j/n),1]}\right)^{(2)}\right]\Big]\Big|\\\overset{(4.4)\text{C)}}{\leq}&\frac{n^{-3/2}}{6}\|g_n\|_{M^0}\mathbb{E}|X_j|^3\end{aligned} \tag{5.3}$$

because, clearly, $\|\mathbb{1}_{[S^{-1}(j/n),1]}\| = 1$. Also, in the last inequality we have used the fact that X_j is independent of $\mathbf{Y}_n^j$. We can now sum (5.3) over $j = 1, 2, ..., \lfloor nS(1) \rfloor$ and use the fact that X_j's are independent of $\mathbf{Y}_n^j$'s and that X_j's have mean 0 and variance 1 to obtain:

$$\left|\mathbb{E}Df_n(\mathbf{Y}_n)[\mathbf{Y}_n] - n^{-1}\sum_{j=1}^{\lfloor nS(1)\rfloor} D^2 f_n(\mathbf{Y}_n^j)\left[\left(\mathbb{1}_{[S^{-1}(j/n),1]}\right)^{(2)}\right]\right| \leq \frac{n^{-1/2}}{6} S(1)\|g\|_{M^0}\mathbb{E}\left|X_1\right|^3.$$

We notice that for $\mathcal{A}_n$ defined in Proposition 4.1, using Remark 4.3, we obtain:

$$\begin{aligned}
&|\mathbb{E}\mathcal{A}_n f_n(\mathbf{Y}_n)| = \left|\mathbb{E}Df_n(\mathbf{Y}_n)[\mathbf{Y}_n] - \mathbb{E}D^2 f_n(\mathbf{Y}_n)\left[\mathbf{A}_n^{(2)}\right]\right| \\
\leq& \left|\mathbb{E}Df_n(\mathbf{Y}_n)[\mathbf{Y}_n] - n^{-1}\sum_{j=1}^{\lfloor nS(1)\rfloor}\mathbb{E}D^2 f_n(Y_n^j)\left[\left(\mathbb{1}_{[S^{-1}(j/n),1]}\right)^{(2)}\right]\right| \\
&+ n^{-1}\left|\sum_{j=1}^{\lfloor nS(1)\rfloor}\mathbb{E}\left\{D^2 f_n(\mathbf{Y}_n)\left[\left(\mathbb{1}_{[S^{-1}(j/n),1]}\right)^{(2)}\right] - D^2 f_n(\mathbf{Y}_n^j)\left[\left(\mathbb{1}_{[S^{-1}(j/n),1]}\right)^{(2)}\right]\right\}\right| \\
\leq& \frac{n^{-1/2}}{6}\|g_n\|_{M^0}\mathbb{E}|X_1|^3 \\
&+ n^{-1}\left|\sum_{j=1}^{\lfloor nS(1)\rfloor}\mathbb{E}\left\{D^2 f_n\left(\mathbf{Y}_n^j + n^{-1/2}X_j\mathbb{1}_{[S^{-1}(j/n),1]}\right)\left[\left(\mathbb{1}_{[S^{-1}(j/n),1]}\right)^{(2)}\right]\right.\right. \\
&\left.\left. -\ D^2 f_n(\mathbf{Y}_n^j)\left[\left(\mathbb{1}_{[S^{-1}(j/n),1]}\right)^{(2)}\right]\right\}\right| \\
\overset{(4.4)C)}{\leq}& \frac{n^{-1/2}}{6}\|g_n\|_{M^0}\mathbb{E}|X_1|^3 + n^{-1}\frac{\|g_n\|_{M^0}}{3}\sum_{j=1}^{\lfloor nS(1)\rfloor} n^{-1/2}\mathbb{E}\left\|X_j\mathbb{1}_{[S^{-1}(j/n),1]}\right\| \\
\leq& \frac{n^{-1/2}}{6}S(1)\|g_n\|_{M^0}\mathbb{E}|X_1|^3 + \frac{n^{-1/2}}{3}S(1)\|g_n\|_{M^0}\mathbb{E}\left\|X_1\mathbb{1}_{[S^{-1}(j/n),1]}\right\| \\
\leq& \frac{n^{-1/2}S(1)\|g_n\|_{M^0}}{2}\mathbb{E}|X_1|^3.
\end{aligned}$$

The last inequality follows by Jensen's inequality:

$$\mathbb{E}|X_1| \leq \sqrt{\mathbb{E}|X_1|^2} = 1 = \left(\mathbb{E}|X_1|^2\right)^{3/2} \leq \mathbb{E}|X_1|^3. \tag{5.4}$$

Now, note that this gives:

$$\begin{aligned}
&|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{A}_n)| = |\mathbb{E}g_n(\mathbf{Y}_n)| = |\mathbb{E}\mathcal{A}_n f_n(\mathbf{Y}_n)| \\
\leq& \frac{n^{-1/2}}{2}S(1)\|g_n\|_{M^0}\mathbb{E}|X_1|^3 \\
\leq& \frac{n^{-1/2}}{2}S(1)\left(\|g\|_{M^0} + \mathbb{E}g(\mathbf{A}_n)\right)\mathbb{E}|X_1|^3
\end{aligned}$$

$$\leq n^{-1/2}S(1)\|g\|_{M^0}\mathbb{E}|X_1|^3. \tag{5.5}$$

Combining this with (5.2):

$$|\mathbb{E}g(\mathbf{Y}_n)-\mathbb{E}g(\mathbf{Z})|\leq \frac{\|g\|_{M^0}}{n^{1/2}}\left(\frac{30}{\sqrt{\pi}\log 2}\sqrt{\log(2S(1)n)}+S(1)\mathbb{E}|X_1|^3\right) \tag{5.6}$$

which proves Theorem 3.1. □

5.2. *Proof of Theorem 3.4*

5.2.1. *Comparing $\tilde{\mathbf{Y}}_n$ to a Poisson random walk*

Note that $\left(\mathbf{P}(\lfloor nS^{(n)}(t)\rfloor), t\in[0,1]\right)$ can be expressed in the following way:

$$\mathbf{P}(\lfloor nS^{(n)}(t)\rfloor)-\lfloor nS^{(n)}(t)\rfloor=\sum_{i=1}^{\lfloor nS^{(n)}(t)\rfloor}X_i,$$

where (X_i+1)'s are i.i.d. Poisson(1). Therefore, we can express $(\tilde{\mathbf{Y}}_n(t), t\in[0,1])$ in the following way:

$$\tilde{\mathbf{Y}}_n(t)=n^{-1/2}\left\{\sum_{i=1}^{\lfloor nS^{(n)}(t)\rfloor}X_i+\mathbf{P}\left(nS^{(n)}(t)\right)-\mathbf{P}\left(\lfloor nS^{(n)}(t)\rfloor\right)-\left(nS^{(n)}(t)-\lfloor nS^{(n)}(t)\rfloor\right)\right\}.$$

We also define:

$$\mathbf{Y}_n(t)=n^{-1/2}\sum_{i=1}^{\lfloor nS^{(n)}(t)\rfloor}X_i.$$

Note that $\left|nS^{(n)}(t)-\lfloor nS^{(n)}(t)\rfloor\right|\leq 1$ for all $t\geq 0$. Also, observe that for all $t\geq 0$:

$$\left|\mathbf{P}\left(nS^{(n)}(t)\right)-\mathbf{P}\left(\lfloor nS^{(n)}(t)\rfloor\right)\right|\leq\mathbf{P}\left(\lfloor nS^{(n)}(t)\rfloor+1\right)-\mathbf{P}\left(\lfloor nS^{(n)}(t)\rfloor\right).$$

By the independence of increments of a Poisson process:

$$\mathbb{E}\|\tilde{\mathbf{Y}}_n-\mathbf{Y}_n\|\leq n^{-1/2}\left[1+\mathbb{E}\left[\max_{1\leq i\leq n}\bar{P}_i\right]\right], \tag{5.7}$$

where $\bar{P}_1,\cdots,\bar{P}_n\overset{\text{i.i.d}}{\sim}\mathbf{Poisson}(1)$. Using the trick from [Das11], we note that, by Jensen's inequality applied to function $x\mapsto\exp(x\log\log n)$, which is convex for $n\geq 3$,

$$\begin{aligned}\exp\left(\log\log n\cdot\mathbb{E}\left[\max_{1\leq i\leq n}\bar{P}_i\right]\right)&\leq\mathbb{E}\left[\exp\left(\log\log n\cdot\max_{1\leq i\leq n}\bar{P}_i\right)\right]\\&=\mathbb{E}\left[\max_{1\leq i\leq n}\exp\left(\log\log n\cdot\bar{P}_i\right)\right]\\&\leq n\mathbb{E}\left[\exp\left(\log\log n\cdot\bar{P}_1\right)\right]\\&=n\exp\left(\log n-1\right)\\&\leq e^{-1}n^2.\end{aligned} \tag{5.8}$$

Now, combining (5.7) and (5.8) we obtain:

$$\mathbb{E}\|\tilde{\mathbf{Y}}_n - \mathbf{Y}_n\| \leq n^{-1/2}\left[1 + \frac{-1 + 2\log n}{\log\log n}\right]. \tag{5.9}$$

Then, for every $g \in M^0$:

$$\begin{aligned}|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\tilde{\mathbf{Y}}_n)| \overset{\text{MVT}}{\leq} & \mathbb{E}\left[\sup_{c\in[0,1]}\|Dg((1-c)\tilde{\mathbf{Y}}_n + c\mathbf{Y}_n)\|\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\|\right] \\ \leq & \|g\|_{M^0}\mathbb{E}\left[\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\|\right] \\ \leq & n^{-1/2}\left[1 + \frac{-1 + 2\log n}{\log\log n}\right]. \end{aligned} \tag{5.10}$$

5.2.2. Comparing the scaled Poisson random walk to a scaled Gaussian random walk

Let $\mathbf{A}_n(t) = n^{-1/2}\sum_{i=1}^{\lfloor nS^{(n)}(t)\rfloor} Z_i, t \in [0,1]$ for $Z_i \overset{i.i.d}{\sim} \mathcal{N}(0,1)$. By (5.5),

$$\begin{aligned}|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{A}_n)| &\leq n^{-1/2}S^{(n)}(1)\|g\|_{M^0}\mathbb{E}|X_1|^3 \\ &\leq n^{-1/2}S^{(n)}(1)\|g\|_{M^0}(1 + 2e^{-1}) \end{aligned} \tag{5.11}$$

because $X_1 \overset{\mathcal{D}}{=} \mathbf{P}(1) - 1$.

5.2.3. Accounting for the difference between the time changes

Now let $\tilde{\mathbf{A}}_n(t) = n^{-1/2}\sum_{i=1}^{\lfloor nS(t)\rfloor} Z_i, t \in [0,1]$. Then:

$$\begin{aligned}\mathbb{E}\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\| =& n^{-1/2}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\sum_{i=\lfloor nS(t)\wedge S^{(n)}(t)\rfloor+1}^{\lfloor nS(t)\vee S^{(n)}(t)\rfloor} Z_i\right|\right] \\ =& n^{-1/2}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\sum_{i=1}^{\lfloor nS(t)\vee S^{(n)}(t)\rfloor - (\lfloor nS(t)\wedge S^{(n)}(t)\rfloor+1)} Z_i\right|\right] \\ \overset{Doob, Jensen}{\leq} & 2n^{-1/2}\sqrt{\mathbb{E}\left|\sum_{i=1}^{\sup_{t\in[0,1]}(\lfloor nS(t)\vee S^{(n)}(t)\rfloor - (\lfloor nS(t)\wedge S^{(n)}(t)\rfloor+1))} Z_i\right|^2} \\ \leq & 2\sqrt{\|S - S^{(n)}\|}. \end{aligned} \tag{5.12}$$

Therefore:

$$|\mathbb{E}g(\mathbf{A}_n) - \mathbb{E}g(\tilde{\mathbf{A}}_n)| \overset{\text{MVT}}{\leq} \mathbb{E}\left[\sup_{c\in[0,1]} \|Dg\left((1-c)\tilde{\mathbf{A}}_n + c\mathbf{A}_n\right)\| \|\mathbf{A}_n - \tilde{\mathbf{A}}_n\|\right]$$
$$\leq \|g\|_{M^0}\mathbb{E}\left[\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\|\right]$$
$$\leq 2\sqrt{\|S - S^{(n)}\|}\|g\|_{M^0}. \tag{5.13}$$

5.2.4. Comparing the scaled Gaussian random walk to Brownian Motion

By (5.2) we get for $\mathbf{Z} = \mathbf{B} \circ S$:

$$|\mathbb{E}g(\tilde{\mathbf{A}}_n) - \mathbb{E}g(\mathbf{Z})| \leq \frac{\|g\|_{M^0}}{n^{1/2}}\left(\frac{30}{\sqrt{\pi}\log 2}\sqrt{\log(2S(1)n)}\right). \tag{5.14}$$

Theorem 3.4 now follows from (5.10), (5.11), (5.13), (5.14). □

5.3. *Proof of Theorem 3.7*

The proof is inspired by the material of [BDM18].

Let $g \in M^0$ and $\tau_0^n = \inf\{t > 0 : \overline{\mathbf{L}}_n(t) = 0\}$. By [SW95, Theorem 11.9], $\mathbb{P}[1 > \tau_0^n]$ tends to zero with exponential speed under the assumption that $x \geq \mu - \lambda$ and so

$$\left|\mathbb{E}\left[(g(\mathbf{Y}_n) - g(\mathbf{Z}))\, \mathbb{1}_{\{1>\tau_0^n\}}\right]\right| \leq c\|g\|_{M^0}e^{-n}, \tag{5.15}$$

for some constant c.

On the event that $\{1 \leq \tau_0^n\}$, for any $t \in [0,1]$ we have

$$\mathbf{Y}_n = \mathbf{Y}_n^1 - \mathbf{Y}_n^{-1},$$

where

$$\mathbf{Y}_n^1(t) = \frac{\mathbf{P}_1(n\lambda t) - n\lambda t}{\sqrt{n}}, \quad \mathbf{Y}_n^{-1}(t) = \frac{\mathbf{P}_{-1}(n\mu t) - n\mu t}{\sqrt{n}}, \quad t \in [0,1].$$

Let $\mathbf{B}^1$ and $\mathbf{B}^2$ be i.i.d standard Brownian Motions such that, if

$$\mathbf{Z}^1(t) = \mathbf{B}^1(\lambda t) \quad \text{and} \quad \mathbf{Z}^{-1}(t) = \mathbf{B}^{-1}(\mu t) \quad \text{for } t \in [0,1]$$

then we have $\mathbf{Z} = \mathbf{Z}^1 + \mathbf{Z}^{-1}$. Therefore, for $g^{(1)}(x) = g(x - \mathbf{Y}_n^{-1})$ and $g^{(-1)}(x) =$

$g(\mathbf{Z}^1 - x)$,

$$\begin{aligned}
&\left|\mathbb{E}\left[(g(\mathbf{Y}_n) - g(\mathbf{Z}))\,\mathbb{1}_{\{1\le\tau_0^n\}}\right]\right| \\
\le&\left|\mathbb{E}\left\{\mathbb{E}\left[\left(g(\mathbf{Y}_n^1 - \mathbf{Y}_n^{-1}) - g(\mathbf{Z}^1 - \mathbf{Y}_n^{-1})\right)\mathbb{1}_{\{1\le\tau_0^n\}}\,\middle|\,\mathbf{Y}_n^{-1}\right]\right\}\right| \\
&+\left|\mathbb{E}\left\{\mathbb{E}\left[\left(g(\mathbf{Z}^1 - \mathbf{Y}_n^{-1}) - g(\mathbf{Z}^1 - \mathbf{Z}^{-1})\right)\mathbb{1}_{\{1\le\tau_0^n\}}\,\middle|\,\mathbf{Z}^1\right]\right\}\right| \\
=&\left|\mathbb{E}\left\{\mathbb{E}\left[\left(g^{(1)}\left(\mathbf{Y}_n^1\right) - g^{(1)}\left(\mathbf{Z}^1\right)\right)\mathbb{1}_{\{1\le\tau_0^n\}}\,\middle|\,\mathbf{Y}_n^{-1}\right]\right\}\right| \\
&+\left|\mathbb{E}\left\{\mathbb{E}\left[\left(g^{(-1)}\left(\mathbf{Y}_n^{-1}\right) - g^{(-1)}\left(\mathbf{Z}^{-1}\right)\right)\mathbb{1}_{\{1\le\tau_0^n\}}\,\middle|\,\mathbf{Z}^1\right]\right\}\right| \qquad (5.16)
\end{aligned}$$

It is clear that (almost surely):

$$\|g^{(1)}\|_{M^0} \le \|g\|_{M^0} \quad \text{and} \quad \|g^{(-1)}\|_{M^0} \le \|g\|_{M^0}.$$

Therefore, using Theorem 3.4,

$$\begin{aligned}
\text{A)}\quad &\left|\mathbb{E}\left\{\mathbb{E}\left[\left(g^{(1)}\left(\mathbf{Y}_n^1\right) - g^{(1)}\left(\mathbf{Z}^1\right)\right)\mathbb{1}_{\{1\le\tau_0^n\}}\,\middle|\,\mathbf{Y}_n^{-1}\right]\right\}\right| \\
&\le\frac{\|g\|_{M^0}}{n^{1/2}}\left(\frac{30}{\sqrt{\pi\log 2}}\sqrt{\log(2\lambda n)} + 2\lambda + 1 + \frac{\log 2 + 2\log n}{\log\log(n+2)}\right); \\
\text{B)}\quad &\left|\mathbb{E}\left\{\mathbb{E}\left[\left(g^{(-1)}\left(\mathbf{Y}_n^{-1}\right) - g^{(-1)}\left(\mathbf{Z}^{-1}\right)\right)\mathbb{1}_{\{1\le\tau_0^n\}}\,\middle|\,\mathbf{Z}^1\right]\right\}\right|; \\
&\le\frac{\|g\|_{M^0}}{n^{1/2}}\left(\frac{30}{\sqrt{\pi\log 2}}\sqrt{\log(2\mu n)} + 2\mu + 1 + \frac{\log 2 + 2\log n}{\log\log(n+2)}\right). \qquad (5.17)
\end{aligned}$$

It now follows from (5.16) and (5.17) that

$$\begin{aligned}
&\left|\mathbb{E}\left[(g(\mathbf{Y}_n) - g(\mathbf{Z}))\,\mathbb{1}_{\{1\le\tau_0^n\}}\right]\right| \\
\le&\frac{\|g\|_{M^0}}{n^{1/2}}\left(\frac{30\left(\sqrt{\log(2\lambda n)} + \sqrt{\log(2\mu n)}\right)}{\sqrt{\pi\log 2}} + 2(\lambda+\mu) + 2 + \frac{2\log 2 + 4\log n}{\log\log(n+2)}\right),
\end{aligned}$$

which, together with (5.15), proves the result. □

Remark 5.1. *Continuous-time Markov chains whose jump rates do not depend on the state the chain is in yet only on the size of the jump may be expressed as sums of independent time-changed Poisson processes. Indeed, suppose that for jump sizes $i_1, i_2, \cdots$, jumps of those sizes occur at rates $\lambda_1, \lambda_2, \cdots$, respectively. Then the resulting continuous-time Markov chain may be described in terms of i.i.d rate 1 Poisson processes $\mathbf{P}_1, \mathbf{P}_2, \cdots$ in the following way:*

$$t \mapsto \sum_{k=1}^{\infty} i_k \mathbf{P}_k(\lambda_k t).$$

An analysis similar to that in the proof of Theorem 3.7 may be applied in this case and also whenever the time changes applied to the Poisson processes

$\mathbf{P}_1, \mathbf{P}_2, \cdots$ *are random, yet independent of those Poisson processes. Indeed, Theorem 3.4 together with a Lindeberg-type trick similar to (5.16) and (5.17) above may be used to bound the distance between such a continuous-time chain and a diffusion process.*

Acknowledgements

The author would like to thank Gesine Reinert, Alison Etheridge and Andrew Barbour for many helpful discussions and Gesine Reinert for constructive comments on the early versions of the paper. The author is also grateful to Sebastian Vollmer for pointing out a mistake in an early version of the proof of Proposition 4.2.

References

[Bar90] A.D. Barbour. Stein's Method for Diffusion Approximations. *Probability Theory and Related Fields*, 84:297–322, 1990.

[BDM18] E. Besançon, L. Decreusefond, and P. Moyal. Stein's method for diffusive limit of Markov processes. arXiv:1805.01691, 2018.

[BHJ92] A.D. Barbour, Lars Holst, and Svante Janson. *Poisson Approximation*. Oxford Studies in Probability. Clarendon Press, 1992.

[BJ09] A.D. Barbour and S. Janson. A functional combinatorial central limit theorem. *Electronic Journal of Probability*, 14(81):2352–2370, 2009.

[CD13] L. Coutin and L. Decreusefond. Stein's method for Brownian Approximations. *Communications on Stochastic Analysis*, 7(3):349–372, 2013.

[CGS11] L.H.Y Chen, L. Goldstein, and Q.-M. Shao. *Normal Approximation by Stein's Method*. Probability and Its Applications. Springer Verlag, 2011.

[Das11] G. Dasarathy. A Simple Probability Trick for Bounding the Expected Maximum of n Random Variables. www.cs.cmu.edu/~gautamd/Files/maxGaussians.pdf, 2011. Accessed on 30/11/2016.

[Dev86] L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlar, 1986.

[FN10] M. Fischer and G. Nappo. On the Moments of the Modulus of Continuity of Ito Processes. *Stochastic Analysis and Applications*, 28(1):103–122, 2010.

[Gau12] N. Gautam. *Analysis of Queues: Methods and Applications*. The Operations Research Series. CRC Press, 2012.

[Kas17] M.J. Kasprzak. Multivariate functional approximations with Stein's method of exchangeable pairs. arXiv:1710.09263, 2017.

[Kas18] M.J. Kasprzak. Stein's method for multivariate Brownian approximations of sums under dependence. arXiv:1708.02521, 2018.

[KDV17] M.J. Kasprzak, A. B. Duncan, and S.J. Vollmer. Note on A. Barbour's paper on Stein's method for diffusion approximations. *Electron. Commun. Probab.*, 22(23):1–8, 2017.

[LRS17] C. Ley, G. Reinert, and Y. Swan. Stein's method for comparison of univariate distributions. *Probability Surveys*, 14:1–52, 2017.

[NP12] I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus*. Cambridge tracts in Mathematics. Cambridge University Press, 2012.

[Rob03] P. Robert. *Stochastic Networks and Queues*. Springer-Verlag Berlin Heidelberg, 2003.

[Ros11] N. Ross. Fundamentals of Stein's Method. *Probability Surveys*, 8:210–293, 2011.

[Ste72] Ch. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, 2:583–602, 1972.

[SW95] A. Shwarz and A. Weiss. *Large Deviations for Performance Analysis: Queues, Communication and Computing*. Chapman and Hall/CRC, London, 1995.

[Swa16] Y. Swan. A gateway to Stein's Method. https://sites.google.com/site/steinsmethod/home, 2016. Accessed on 19/05/2016.

Appendix A: Proof of Proposition 2.2

Note that the proof of Proposition 3.1 of [BJ09] readily applies in this case up to and excluding (3.4) and it suffices to prove that $\liminf_{n\to\infty} \mathbb{P}[\mathbf{Y}_n \in B] \geq \mathbb{P}[\mathbf{Z} \in B]$ for all sets B of the form $B = \bigcap_{1\leq l\leq L} B_l$, where $B_l = \{w \in D : \|w - s_l\| < \gamma_l\}$, $s_l \in C([0,1],\mathbb{R})$ and γ_l is such that $\mathbb{P}[\mathbf{Z} \in \partial B_l] = 0$.

We will condition on the fact that the minimum holding time (interval of constancy of $\mathbf{Y}_n$) is of length greater that $r_n = \lambda_n^{-3}$. It follows from Theorems 2.1 and 2.2 of Chapter 5 in [Dev86] that if we condition on the number of holding times being equal to i, their lengths are distributed uniformly over the simplex $A_i = \left\{(x_1, ..., x_i) : x_j \geq 0, \sum_{j=1}^i x_j \leq 1\right\}$. Note that the probability of the minimum of them being greater or equal to r_n is $(1-ir_n)^i$ if $i \leq r_n$ and 0 otherwise. This is because $\text{Vol}(A_i) = \frac{1}{i!}$ and $\text{Vol}\left(\left\{(x_1, ..., x_i) : x_j \geq r_n, \sum_{j=1}^i x_j \leq 1\right\}\right) = \frac{(1-ir_n)^i}{i!}$. Therefore:

$$\begin{aligned}
&\mathbb{P}\left[\text{minimal waiting time} \geq r_n\right]\\
&= \sum_{i=1}^{\infty} \mathbb{P}\left[\text{minimal waiting time} \geq r_n | \#\text{waiting times} = i\right] \mathbb{P}\left[\#\text{waiting times} = i\right]\\
&= \sum_{i=1}^{\lfloor \lambda_n^3 \rfloor} (1 - i\lambda_n^{-3})^i e^{-\lambda_n} \frac{(\lambda_n)^{i-1}}{(i-1)!} \xrightarrow{n\to\infty} 1. \qquad (5.18)
\end{aligned}$$

To see this note the following:

$$\text{A)} \sum_{i=\lceil \lambda_n^{5/4} \rceil}^{\lfloor \lambda_n^3 \rfloor} (1 - i\lambda_n^{-3})^i e^{-\lambda_n} \frac{(\lambda_n)^{i-1}}{(i-1)!} \leq e^{-\lambda_n} \left(\lambda_n^3 - \lambda_n^{5/4}\right) \left(1 - \lambda_n^{-7/4}\right)^{\lambda_n^{5/4}} \frac{\lambda_n^{\lambda_n^{5/4}-1}}{\left(\lceil \lambda_n^{5/4} \rceil - 1\right)!}$$

$$\leq \frac{\lambda_n^2 \left\lceil \lambda_n^{5/4} \right\rceil}{e^{\lambda_n}} \cdot \lambda_n^{-\frac{1}{8}\lceil \lambda_n^{5/4} \rceil + \frac{9}{8}\lceil \lambda_n^{9/8} \rceil} \xrightarrow{n\to\infty} 0$$

$$\text{B)} \sum_{i=1}^{\lceil \lambda_n^{5/4} \rceil - 1} (1 - i\lambda_n^{-3})^i e^{-\lambda_n} \frac{(\lambda_n)^{i-1}}{(i-1)!} \geq (1 - \lambda_n^{-7/4})^{\lceil \lambda_n^{5/4} \rceil} e^{-\lambda_n} \sum_{i=1}^{\lceil \lambda_n^{5/4} \rceil - 1} \frac{(\lambda_n)^{i-1}}{(i-1)!} \xrightarrow{n\to\infty} 1,$$

where the convergence in B) holds since $(1-\lambda_n^{-7/4})^{\lceil \lambda_n^{5/4} \rceil} \to 1$, $e^{-\lambda_n} \sum_{i=1}^{\infty} \frac{(\lambda_n)^{i-1}}{(i-1)!} = 1$ and:

$$e^{-\lambda_n} \sum_{i=\lceil \lambda_n^{5/4} \rceil}^{\infty} \frac{(\lambda_n)^{i-1}}{(i-1)!} \leq e^{-\lambda_n} \frac{\lambda_n^{\lceil \lambda_n^{5/4} \rceil}}{\lceil \lambda_n^{5/4} \rceil !} \cdot \frac{\lceil \lambda_n^{5/4} \rceil + 1}{\lceil \lambda_n^{5/4} \rceil + 1 - \lambda_n} \xrightarrow{n\to\infty} 0$$

for instance, by Proposition A.2.3(ii) of [BHJ92]. Furthermore, note that for $g_{l,n}^*$ defined by (3.6) in [BJ09]:

$$\begin{aligned}
&\liminf_{n\to\infty} \mathbb{E}\left[\prod_{l=1}^{L} g_{l,n}^*(\mathbf{Y}_n)\right] \\
=&\liminf_{n\to\infty} \mathbb{E}\left[\prod_{l=1}^{L} g_{l,n}^*(\mathbf{Y}_n) \middle| \text{minimal waiting time} \geq r_n\right] \mathbb{P}\left[\text{minimal waiting time} \geq r_n\right] \\
&+ \liminf_{n\to\infty} \mathbb{E}\left[\prod_{l=1}^{L} g_{l,n}^*(\mathbf{Y}_n) \middle| \text{minimal waiting time} < r_n\right] \mathbb{P}\left[\text{minimal waiting time} < r_n\right] \\
=&\liminf_{n\to\infty} \mathbb{E}\left[\prod_{l=1}^{L} g_{l,n}^*(\mathbf{Y}_n) \middle| \text{minimal waiting time} \geq r_n\right] \mathbb{P}\left[\text{minimal waiting time} \geq r_n\right]
\end{aligned} \tag{5.19}$$

because:

$$\begin{aligned}
0 \leq &\liminf_{n\to\infty} \mathbb{E}\left[\prod_{l=1}^{L} g_{l,n}^*(\mathbf{Y}_n) \middle| \text{minimal waiting time} < r_n\right] \mathbb{P}\left[\text{minimal waiting time} < r_n\right] \\
\leq &\liminf_{n\to\infty} \mathbb{P}\left[\text{minimal waiting time} < r_n\right] \overset{(5.18)}{=} 0.
\end{aligned}$$

Following the same steps as in [BJ09], we obtain:

$$\begin{aligned}
&\liminf_{n\to\infty}\mathbb{P}\left[\mathbf{Y}_n\in B\right]\geq\liminf_{n\to\infty}\mathbb{P}\left[\mathbf{Y}_n\in B\text{ and minimal waiting time}\geq r_n\right]\\
&\geq\liminf_{n\to\infty}\mathbb{E}\left[\prod_{l=1}^{L}g^*_{l,n}(\mathbf{Y}_n)\middle|\text{minimal waiting time}\geq r_n\right]\mathbb{P}\left[\text{minimal waiting time}\geq r_n\right]\\
&\overset{(2.1),(5.19)}{\geq}\liminf_{n\to\infty}\left\{\mathbb{E}\left[\prod_{l=1}^{L}g^*_{l,n}(\mathbf{Z}_n)\right]-C\tau_n\left\|\prod_{l=1}^{L}g^*_{l,n}\right\|_{M^0}\right\}\\
&\geq\liminf_{n\to\infty}\left\{\mathbb{E}\left[\prod_{l=1}^{L}g^*_{l,n}(\mathbf{Z}_n)\right]-C''\tau_n{p_n}^2(\epsilon\gamma)^{-2}\eta_n^{-3}\right\}\\
&\overset{\text{Fatou}}{\geq}\mathbb{E}\left[\liminf_{n\to\infty}\prod_{l=1}^{L}g^*_{l,n}(\mathbf{Z}_n)\right]\geq\mathbb{P}\left[\bigcap_{1\leq l\leq L}\left(\|\mathbf{Z}-s_l\|<\gamma_l(1-\theta)\right)\right].
\end{aligned}$$

□

arXiv: 1708.02521

Stein's method for multivariate Brownian approximations of sums under dependence

Mikołaj J. Kasprzak,

Mikołaj Kasprzak
Department of Statistics
University of Oxford
24-29 St Giles'
Oxford, OX1 3LB
United Kingdom
e-mail: kasprzak@stats.ox.ac.uk

Abstract: We use Stein's method to obtain a bound on the distance between scaled p-dimensional random walks and a p-dimensional (correlated) Brownian Motion. We consider dependence schemes including those in which the summands in scaled sums are weakly dependent and their p components are strongly correlated. As an example application, we prove a functional limit theorem for exceedances in an m-scans process, together with a bound on the rate of convergence. We also find a bound on the rate of convergence of scaled U-statistics to Brownian Motion, representing an example of a sum of strongly dependent terms.

MSC 2010 subject classifications: Primary 60B10, 60F17; secondary 60B12, 60J65, 60E05, 60E15.
Keywords and phrases: Stein's method, functional convergence, Brownian Motion, scaled random walk, u-statistics.

1. Introduction

In the seminal paper [Bar90], Barbour addressed the problem of providing bounds on the rate of convergence in functional limit results (or invariance principles as they are often called in the literature). He observed that the celebrated Stein's method, first introduced in [Ste72] as a tool for proving the Central Limit Theorem, may also be used in the setup of the *Functional* Central Limit Theorem. This theorem, whose early versions are attributed to Donsker [Don51], says that for a sequence of i.i.d. real random variables $(X_n)_{n=1}^{\infty}$ with mean zero and unit variance, the random process

$$\mathbf{Y}_n(t) = n^{-1/2}\sum_{i=1}^{\lfloor nt \rfloor} X_i, \quad t \in [0,1] \tag{1.1}$$

converges in distribution to the standard Brownian Motion with respect to the Skorokhod topology.

Through a careful and technical adaptation of Stein's method to the framework of a Brownian-Motion approximation and a subsequent repetitive use of Taylor's theorem, Barbour [Bar90] proved a powerful estimate on a distance between the law of $\mathbf{Y}_n$ in (1.1) and the Wiener measure. Specifically, he considered test functions g acting on the Skorokhod space $D([0,1],\mathbb{R})$ of càdlàg real-valued maps on $[0,1]$, such that g takes values in the reals, does not grow faster than a cubic, is twice Fréchet differentiable and its second derivative is Lipschitz. Denoting by $\mathbf{Z}$ the Brownian Motion on $[0,1]$ and adopting the notation of (1.1), his result says that

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{Z})| \leq C_g \frac{\mathbb{E}|X_1|^3 + \sqrt{\log n}}{\sqrt{n}},$$

where C_g is a constant, independent of n, yet depending on the (carefully defined) *smoothness properties* of g. Among the applications and extensions considered by Barbour are an analysis of the empirical distribution function of i.i.d. random variables and the Wald-Wolfowitz theorem often used to construct tests in non-parametric statistics [WW40].

Our aim in this paper is to extend the results of [Bar90] to approximations of scaled sums of univariate and *multivariate* random variables with different *dependence structures* by univariate and *multivariate* Wiener processes.

1.1. Motivation

Functional limit results play an important role in applied fields. Researchers often choose to model discrete phenomena with continuous processes arising as scaling limits of discrete ones. The reason is that those scaling limits may be studied using stochastic analysis and are more robust to changes in local details. Questions about the rate of convergence in functional limit results are equivalent to ones about the error those researchers make when doing so. Obtaining bounds on a certain distance between the scaled discrete and the limiting continuous processes provides a way of quantifying this error.

Our motivation in this paper comes from the desire to fill in a gap in the theory but we are also motivated by examples related to applications.

One of those, Example 3.4 of this paper, considers *exceedances of the m-scans process*. For a sequence of i.i.d. random variables $X_1, X_2, \cdots$, the one-dimensional m-scans process is given by $R_i = \sum_{k=0}^{m-1} X_{i+k}$. The number of its exceedances of a real number a is given by

$$Y = \sum_{i=1}^{n} \mathbb{1}[R_i > a].$$

As noted in [CGS11, Example 9.2], this statistic has been studied by many authors, including [GNW01] and [Nau82]. It is of high importance in many areas of applied statistics and has been used, for instance, to evaluate the significance of observed inhomogeneities in the distribution of markers along the length of long DNA sequences (see [DK92, KB92]). Y may be normalized and centralized and then shown to converge in distribution to the standard normal law. Berry-Esseen bounds on the rate of this convergence have been found in [DR96, Theorem 4.1] and [CGS11, Example 9.2]. We are interested in studying the functional convergence of a multidimensional version of Y.

Another example concerns *U-statistics* and is treated in Theorem 3.9 of this paper. U-statistics are defined to be random variables of the form:

$$S_n^k(h) = \sum_{1\leq i_1<...<i_k\leq n} h(X_{i_1}, ..., X_{i_k}), \quad n \geq 1$$

for a symmetric real (or complex) function h on $\mathcal{S}^k$ (where $\mathcal{S}$ is some measurable space) and a sequence of i.i.d. random variables $(X_i)_{i\geq 1}$ taking values in $\mathcal{S}$. Because of their appealing properties, they are central objects in the field of Mathematical Statistics, as described in [KJ88] and many commonly used statistics can be expressed in terms of certain U-statistics or approximated by them. They also appear in decompositions of more general statistics into sums of terms of a simpler form (see, e.g. [Ser80, Chapter 6] or [RV80] and [Vit84]) and play an important role in the study of random fields (see, e.g. [Chr87, Chapter 4]). The appealing properties of *non-degenerate U-statistics*, i.e. those such that, for

$$w(x) = \mathbb{E}h(x, X_1, \cdots, X_{k-1}),$$

$0 < \text{Var}[w(X_1)] < \infty$, include their asymptotic behaviour. It can be described by a Strong Law of Large Numbers ([Hoe61]), a central limit theorem ([Hoe48]) or the functional central limit theorem (e.g. [Jan97, Chapter XI]), which will be studied in this paper. Other interesting results include those connected to large deviations for U-statistics (see [EL99]), Berry-Esseen-type bounds (see [CS07]) and other bounds on the speed of convergence in the U-statistic CLT (see [RR97]). *Degenerate U-statistics* have also received much attention in the recent years with [DP17] providing bounds on the speed of convergence in de Jong's theorem [dJ90] and proving its multidimensional version.

Our theoretical motivation is expressed in Proposition 3.5 of this paper. It seems natural to ask whether techniques similar to those of [Bar90] may be used to study a process of the form

$$t \mapsto n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} X_i, \quad t \in [0, 1] \tag{1.2}$$

where $\{X_i : i = 1, \cdots, n\}$ is a collection of i.i.d. random vectors in $\mathbb{R}^p$ for $p > 1$ with a given covariance matrix Σ. Interesting questions arising include those about the rate of convergence of the process in (1.2) to the correlated p-dimensional Brownian Motion created from a standard Brownian Motion $\mathbf{B}$ by premultiplying it by $\Sigma^{1/2}$. In this context, the role played by Σ in the quality of this approximation seems worth paying attention to.

1.2. Contribution of the paper

The main achievements of the paper are the following:

(a) A very general result providing a bound on the distance between a process of the form

$$\mathbf{Y}_n(t) = \left(\sum_{i=1}^{\lambda_1} X_{i,1}J_{i,1}(t), \cdots, \sum_{i=1}^{\lambda_p} X_{i,p}J_{i,p}(t)\right), \quad t \in [0,1],$$

where:

- the collection of vectors $X_i = (X_{i,1}, \cdots, X_{i,p})$ for $i = 1, \cdots, n$ is allowed to be *dependent* and those vectors themselves are allowed to have non-identity covariance matrices
- the collection of (possibly random) functions

$$\{J_{i,k} \in D([0,1], \mathbb{R}) : i = 1, \cdots n, k = 1, \cdots, p\}$$

is independent of the collection of vectors $(X_i)_{i=1}^n$ from the previous point;
- the numbers λ_j are such that $\lambda_j \le n$

and a correlated p-dimensional Brownian Motion. The bound is presented in Theorem 3.1 and provides a substantial extension of the result of [Bar90], which bounds the rate of convergence in the classical, one-dimensional Donsker's invariance principle.

(b) A novel functional central limit theorem invloving the number of exceedances in the multidimensional m-scans process, together with a bound on the rate of convergence, presented in Example 3.4.

(c) A novel bound on the rate of convergence in the functional central limit theorem for non-degenerate, bivariate U-statistics (for a classical proof of the theorem see, for instance, [Hal79]), which is presented in Theorem 3.9.

(d) A technical result, presented in Proposition 2.3, showing that our bounds' converging to zero implies weak convergence of the underlying processes with respect to the Skorokhod and uniform topologies. This result is a direct extension of [BJ09, Proposition 3.1] to the multidimensional setting.

We provide explicit values for all the constants appearing in our bounds. To our best knowledge, none of the authors who have considered functional approximations with Stein's method so far has done so. We do it as we hope that this will make our results more powerful when used in applications.

The technique which is central in obtaining all the bounds is Stein's method.

1.3. Stein's method for distributional approximation

In [Ste72] it is observed that a random variable Z has standard normal law if and only if $\mathbb{E}Zf(Z) = \mathbb{E}f'(Z)$ for all smooth functions f. Therefore, if, for a random variable W with mean 0 and variance 1, $\mathbb{E}f'(W) - \mathbb{E}Wf(W)$ is close to zero for a large class of functions f, then the law of W should be approximately Gaussian. This leads to a method of bounding the speed of convergence to the normal distribution. Instead of evaluating $|\mathbb{E}h(W) - \mathbb{E}h(Z)|$ directly for a given function h, one can first find an $f = f_h$ solving the following *Stein equation*:

$$f'(w) - wf(w) = h(w) - \mathbb{E}h(Z)$$

and then find a bound on $|\mathbb{E}f'(W) - \mathbb{E}Wf(W)|$. This approach, called *Stein's method*, often turns out to be surprisingly easy and has also proved to be useful for approximations by distributions other than normal.

The aim of the generalised version of Stein's method is to find a bound of the quantity $|\mathbb{E}_{\nu_n}h - \mathbb{E}_\mu h|$, where μ is the target (known) distribution, ν_n is the approximating law and h is chosen from a suitable class of real-valued test functions $\mathcal{H}$. The procedure can be described in terms of three steps. First, an operator $\mathcal{A}$ acting on a class of real-valued functions is sought, such that

$$(\forall f \in \text{Domain}(\mathcal{A}) \quad \mathbb{E}_\nu \mathcal{A}f = 0) \quad \Longleftrightarrow \quad \nu = \mu,$$

where μ is our target distribution. Then, for a given function $h \in \mathcal{H}$, the Stein equation

$$\mathcal{A}f = h - \mathbb{E}_\mu h$$

is solved. Finally, using properties of the solution and various mathematical tools (among which the most popular are Taylor's expansions in the continuous case, Malliavin calculus, as described in [NP12], and coupling methods), an explicit bound is sought for the quantity $|\mathbb{E}_{\nu_n}\mathcal{A}f_h|$.

An accessible account of the method can be found, for example, in the surveys [LRS17] and [Ros11] as well as the books [BHJ92] and [CGS11], which treat the cases of Poisson and normal approximation, respectively, in detail. [Swa16] is a database of information and publications connected to Stein's method.

Approximations by laws of diffusion processes have not been covered in the Stein's method literature very widely, with the notable exceptions of [Bar90, BJ09, Shi11, CD13] and recently [BDM18, Kas17a, Kas17b]. Our aim in this paper is to develop it in a direction not previously explored by other authors while completely natural given the direction in which the finite-dimensional Stein's method literature has evolved.

1.4. Structure of the paper

In Section 2 we define the spaces of test functions we will be working with and the corresponding norms which will appear in the bounds. We also present Proposition 2.3 giving circumstances under which the bounds obtained later in the paper converging to zero imply weak convergence of the considered probability distributions. Section 3 gives statements of the main results of the paper, mentioned above. Section 4 contains all the proofs preceded by finding the Stein equation for approximation by the law of interest, solving it and examining properties of the solutions. In the appendix we present the proof of the aforementioned Proposition 2.3.

2. Notation and spaces M, M^1, M^2 and M^0

The following notation is used throughout the paper. For a function w defined on the interval $[0,1]$ and taking values in a Euclidean space, we define

$$\|w\| = \sup_{t\in[0,1]} |w(t)|,$$

where $|\cdot|$ denotes the Euclidean norm. We also let $D^p = D([0,1],\mathbb{R}^p)$ be the Skorokhod space of all càdlàg functions on $[0,1]$ taking values in $\mathbb{R}^p$. In the sequel, for $i = 1,\cdots,p$, e_i will denote the ith unit vector of the canonical basis of $\mathbb{R}^p$ and the ith component of $x \in \mathbb{R}^p$ will be represented by $x^{(i)}$, i.e. $x = \left(x^{(1)},\cdots,x^{(p)}\right)$.

Let $p \in \mathbb{N}$. Let us define:

$$\|f\|_L := \sup_{w\in D^p} \frac{|f(w)|}{1+\|w\|^3},$$

and let L be the Banach space of continuous functions $f : D^p \to \mathbb{R}$ such that $\|f\|_L < \infty$. Following [Bar90], we now let $M \subset L$ consist of the twice Fréchet differentiable functions f, such that:

$$\|D^2f(w+h) - D^2f(w)\| \leq k_f\|h\|, \tag{2.1}$$

for some constant k_f, uniformly in $w, h \in D^p$. By D^kf we mean the k-th Fréchet derivative of f and the norm of k-linear form B on L is defined to be

$$\|B\| = \sup_{\{h:\|h\|=1\}} |B[h,...,h]|.$$

Note the following lemma, which can be proved in an analogous way to that used to show (2.6) and (2.7) of [Bar90]. We omit the proof here.

Lemma 2.1. *For every $f \in M$, let:*

$$\begin{aligned}\|f\|_M := &\sup_{w\in D^p} \frac{|f(w)|}{1+\|w\|^3} + \sup_{w\in D^p} \frac{\|Df(w)\|}{1+\|w\|^2} + \sup_{w\in D^p} \frac{\|D^2f(w)\|}{1+\|w\|}\\ &+ \sup_{w,h\in D^p} \frac{\|D^2f(w+h)-D^2f(w)\|}{\|h\|}.\end{aligned}$$

Then, for all $f \in M$, we have $\|f\|_M < \infty$.

For future reference, we let $M^1 \subset M$ be the class of functionals $g \in M$ such that:

$$\begin{aligned}\|g\|_{M^1} := &\sup_{w\in D^p} \frac{|g(w)|}{1+\|w\|^3} + \sup_{w\in D^p} \|Dg(w)\| + \sup_{w\in D^p} \|D^2g(w)\|\\ &+ \sup_{w,h\in D^p} \frac{\|D^2g(w+h)-D^2g(w)\|}{\|h\|} < \infty.\end{aligned} \tag{2.2}$$

and $M^2 \subset M$ be the class of functionals $g \in M$ such that:

$$\|g\|_{M^2} := \sup_{w\in D^p} \frac{|g(w)|}{1+\|w\|^3} + \sup_{w\in D^p} \frac{\|Dg(w)\|}{1+\|w\|} + \sup_{w\in D^p} \frac{\|D^2g(w)\|}{1+\|w\|} + \sup_{w,h\in D^p} \frac{\|D^2g(w+h)-D^2g(w)\|}{\|h\|} < \infty. \tag{2.3}$$

We also let M^0 be the class of functionals $g \in M$ such that:

$$\|g\|_{M^0} := \sup_{w\in D^p} |g(w)| + \sup_{w\in D^p} \|Dg(w)\| + \sup_{w\in D^p} \|D^2g(w)\| + \sup_{w,h\in D^p} \frac{\|D^2g(w+h)-D^2g(w)\|}{\|h\|} < \infty.$$

We note that $M^0 \subset M^1 \subset M^2 \subset M$. We shall refer to those different classes of functions in the results presented in the remainder of this paper. In each case we aim to obtain our bounds for the largest possible class, yet it is not always possible to do so for class M or even M^2. Hence, the introduction of the above presented restrictions of M is necessary for a recovery of the full strength of our results.

The next proposition is a p-dimensional version of [BJ09, Proposition 3.1] and shows conditions, under which convergence of the sequence of expectations of a functional g under the approximating measures to the expectation of g under the target measure for all $g \in M^0$ implies weak convergence of the measures of interest. The proposition will be later used to conclude weak convergence from bounds derived in the theorems of the next section. Its proof can be found in the appendix.

Definition 2.2. *$Y \in D([0,1],\mathbb{R}^p)$ is piecewise constant if $[0,1]$ can be divided into intervals of constancy $[a_k, a_{k+1})$ such that the Euclidean norm of $(Y(t_1) - Y(t_2))$ is equal to 0 for all $t_1, t_2 \in [a_k, a_{k+1})$.*

Proposition 2.3. *Suppose that, for each $n \geq 1$, the random element $\mathbf{Y}_n$ of D^p is piecewise constant and let $r_n > 0$ be such that the intervals of constancy are of length at least r_n. Let $(\mathbf{Z}_n)_{n\geq 1}$ be random elements of D^p converging weakly in D^p, with respect to the Skorokhod topology, to a random element $\mathbf{Z} \in C([0,1],\mathbb{R}^p) \subset D^p$. If there exists τ_n such that $\tau_n \log^2(1/r_n) \xrightarrow{n\to\infty} 0$ and*

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{Z}_n)| \leq C\tau_n \|g\|_{M^0} \tag{2.4}$$

for each $g \in M^0$ then $\mathbf{Y}_n \Rightarrow Z$ (converges weakly) in D^p, in both the uniform and the Skorokhod topology.

Remark 2.4. *The formulation of Proposition 2.3 is almost identical to that of [BJ09, Proposition 3.1] with the only difference being that $\mathbf{Y}_n$ and $\mathbf{Z}_n$ are allowed to be p-dimensional for $p > 1$. For completeness, the appendix contains a more detailed proof than the one presented in [BJ09], which may be used by the reader to derive extensions or other versions of the result.*

3. Main results

Theorem 3.1 below treats the case in which the summands in the scaled sum are locally dependent and their components are (strongly) dependent. The error bound in the approximation by a correlated Brownian Motion is obtained for functions in M^1 of (2.2).

Theorem 3.1 (Dependent components and locally dependent summands)**.** *Let n and p be non-negative integers. Consider an array of mean-zero random variables*

$$\{X_{i,j} : i = 1, ..., n, j = 1, ..., p\},$$

with a positive definite covariance matrix. Let

(a) $\lambda_j \leq n$, *for* $j = 1, \cdots p$, *be deterministic positive integers;*
(b) $\mathbb{A}_i \subset \{1, 2, ..., n\}$, *for* $i = 1, \cdots n$ *be a set such that* $X_i = (X_{i,1}, \cdots, X_{i,p})$ *is independent of* $\{X_j : j \in \mathbb{A}_i^c\}$;
(c) $J_{i,k} \in D([0, 1], \mathbb{R})$ *for* $i = 1, \cdots n$ *and* $k = 1, \cdots, p$, *be (possibly random) functions, independent of the* $X_{i,k}$*'s.*

Assume that:

(1)

$$K := \sup_{\substack{i,j \in \{1,\cdots,n\},\\ k,l \in \{1,\cdots,p\}}} \mathbb{E}\left[\|J_{i,k}\|\|J_{j,l}\|\right] < \infty \tag{3.1}$$

(2)

$$\sup_{\substack{i_1,i_2,i_3 \in \{1,\cdots,n\}\\ k_1,k_2,k_3 \in \{1,\cdots,p\}}} \mathbb{E}\left[\|J_{i_1,k_1}\|\|J_{i_2,k_2}\|\|J_{i_3,k_3}\|\right] < \infty.$$

Let

$$\mathbf{Y}_n(t) = \left(\sum_{i=1}^{\lambda_1} X_{i,1} J_{i,1}(t), \cdots, \sum_{i=1}^{\lambda_p} X_{i,p} J_{i,p}(t)\right), \quad t \in [0, 1].$$

Furthermore, for a standard p-dimensional Brownian Motion $\mathbf{B}$ and a positive definite covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$, let $\mathbf{Z} = \Sigma^{1/2}\mathbf{B}$. Then, for any $g \in M^1$, as defined by (2.2):

$$|\mathbb{E}g(\mathbf{Y_n}) - \mathbb{E}g(\mathbf{Z})| \leq \|g\|_{M^1} \sum_{i=1}^{7} \epsilon_i,$$

where:

$$\epsilon_1 = \frac{1}{6}\sum_{i=1}^n \mathbb{E}\left\{\left(\sum_{k,l,m=1}^p \left[(X_{i,k})^2 \|J_{i,k}\|^2 \mathbb{1}_{[1,\lambda_k]}(i)\left(\sum_{j\in\mathbb{A}_i} X_{j,l}\|J_{j,l}\|\mathbb{1}_{[1,\lambda_l]}(j)\right)^2 \cdot \left(\sum_{j\in\mathbb{A}_i} X_{j,m}\|J_{j,m}\|\mathbb{1}_{[1,\lambda_m]}(j)\right)^2\right]\right)^{1/2}\right\};$$

$$\epsilon_2 = \frac{1}{3}\sum_{i=1}^n\sum_{j\in\mathbb{A}_i}\sum_{k,l=1}^p \mathbb{E}\left\{\left[\sum_{m=1}^p\left(X_{i,k}\|J_{i,k}\| X_{j,l}\|J_{j,l}\|\mathbb{1}_{[1,\lambda_k]}(i)\mathbb{1}_{[1,\lambda_l]}(j) \cdot \sum_{r\in\mathbb{A}_j\cap\mathbb{A}_i^c} X_{r,m}\|J_{r,m,n}\|\mathbb{1}_{[1,\lambda_m]}(r)\right)^2\right]^{1/2}\right\};$$

$$\epsilon_3 = \frac{1}{3}\sum_{i=1}^n\sum_{j\in\mathbb{A}_i}\sum_{k,l=1}^p\left\{\mathbb{E}\left[|X_{i,k}X_{j,l}|\right]\mathbb{1}_{[1,\lambda_k]}(i)\mathbb{1}_{[1,\lambda_l]}(j) \cdot \mathbb{E}\left[\|J_{i,k}\|\,\|J_{j,l}\|\sqrt{\sum_{m=1}^p\left(\sum_{r\in\mathbb{A}_i\cup\mathbb{A}_j} X_{r,m}\|J_{r,m}\|\mathbb{1}_{[1,\lambda_m]}(r)\right)^2}\right]\right\};$$

$$\epsilon_4 = \frac{K}{2}\sum_{i=1}^n\sum_{k,l=1}^p\left|\frac{\Sigma_{k,l}}{\sqrt{\lambda_k\lambda_l}} - \mathbb{E}[X_{i,k}X_{i,l}]\right|;$$

$$\epsilon_5 = \frac{K}{2}\sum_{i=1}^n\sum_{j\in\mathbb{A}_i\setminus\{i\}}\sum_{k,l=1}^p\left|\mathbb{E}[X_{i,k}X_{j,l}]\right|;$$

$$\epsilon_6 = \frac{6\sqrt{5}}{\sqrt{2\log 2}}\left(\sum_{i=1}^p\frac{\log(2\lambda_i)}{\lambda_i}\right)^{1/2}\left(\sum_{i=1}^p\Sigma_{i,i}\right)^{1/2};$$

$$\epsilon_7 = \sum_{k=1}^p\sum_{i=1}^{\lambda_k}\frac{\sqrt{\Sigma_{ii}}}{\sqrt{\lambda_k}}\mathbb{E}\left\|J_{i,k} - \mathbb{1}_{[i/\lambda_k,1]}\right\|.$$

Remark 3.2 (Relevance of terms in the bound)**.**

(a) Terms $\epsilon_1,\epsilon_2,\epsilon_3$ correspond to a Berry-Esseen-type bound involving third moments of the summands, and also account for local dependence between the summands;

(b) Terms ϵ_4 and ϵ_5 involve a variance estimation with the latter corresponding to the off-diagonal terms of the covariance matrix of the summands, accounting for the dependence;

(c) Term ϵ_6 comes from estimates on the moments of the Brownian modulus of continuity and accounts for the transition from the Skorokhod space to the Wiener space of continuous functions

(d) Term ϵ_7 describes the randomness of the functions $J_{i,k,n}$ and their distance from indicators $\mathbb{1}_{[i/\lambda_k,1]}$.

Remark 3.3 (Independent summands)**.** *If the summands are independent in Theorem 3.1, i.e. $\mathbb{A}_i = \{i\}$ for all i, then ϵ_2 and ϵ_5 disappear from the bound and ϵ_1 and ϵ_3 become simpler. The new bound takes the following form*

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{Z})| \leq \|g\|_{M^1}(\epsilon_1 + \epsilon_3 + \epsilon_4 + \epsilon_6 + \epsilon_7),$$

where:

$$\epsilon_1 = \frac{1}{6}\sum_{i=1}^n \mathbb{E}\left\{\left[\sum_{k,l,m=1}^p \left(X_{i,k}X_{i,l}X_{i,m}\|J_{i,k}\|\,\|J_{i,l}\|\,\|J_{i,m}\|\mathbb{1}_{[1,\lambda_k]\cap[1,\lambda_l]\cap[1,\lambda_m]}(i)\right)^2\right]^{1/2}\right\};$$

$$\epsilon_3 = \frac{1}{3}\sum_{k,l=1}^p \sum_{i=1}^{\min(\lambda_k,\lambda_l)}\left\{\mathbb{E}\left[|X_{i,k}X_{i,l}|\right]\mathbb{E}\left[\|J_{i,k}\|\,\|J_{i,l}\|\sqrt{\sum_{m=1}^p\left(X_{i,m}\|J_{i,m}\|\mathbb{1}_{[1,\lambda_m]}(i)\right)^2}\right]\right\};$$

$$\epsilon_4 = \frac{K}{2}\sum_{i=1}^n\sum_{k,l=1}^p\left|\frac{\Sigma_{k,l}}{\sqrt{\lambda_k\lambda_l}} - \mathbb{E}[X_{i,k}X_{i,l}]\right|;$$

$$\epsilon_6 = \frac{6\sqrt{5}}{\sqrt{2\log 2}}\left(\sum_{i=1}^p \frac{\log(2\lambda_i)}{\lambda_i}\right)^{1/2}\left(\sum_{i=1}^p \Sigma_{i,i}\right)^{1/2};$$

$$\epsilon_7 = \sum_{k=1}^p\sum_{i=1}^{\lambda_k}\frac{\sqrt{\Sigma_{ii}}}{\sqrt{\lambda_k}}\mathbb{E}\left\|J_{i,k} - \mathbb{1}_{[i/\lambda_k,1]}\right\|.$$

A bound for functions in the larger class M (see Section 2), in the case of independent summands, is obtained in Proposition 3.5.

Example 3.4 (Exceedances of the m-scans process)**.** *Consider an extension of the one-dimensional [CGS11, Example 9.2, p. 254] to the multidimensional and functional setting. For $j = 1, 2, \cdots$, let $V_j = (V_{j,1}, \cdots, V_{j,p})$ be i.i.d. random vectors in $\mathbb{R}^p$. For $k = 1, \cdots, p$ and $i = 1, \cdots, n$ let $R_{i,k} = \sum_{l=0}^{m-1} V_{i+l,k}$ be an m-scans process. Let $a = (a_1, \cdots, a_p) \in \mathbb{R}^p$ and suppose that $n > m$.*

For $k = 1, \cdots, p$, let $\pi_k = \mathbb{P}(R_{1,k} \leq a_k)$ and for $i = 1, \cdots, n$ and $k = 1, \cdots, p$, let

$$X_{i,k} = \frac{1}{n}\left(\sum_{j=1}^n \mathbb{1}[R_{n(i-1)+j,k} \leq a_k]\right) - \pi_k.$$

Extending [DR96, (4.1)], we have that, for $k, l = 1, \cdots, p$,

$$\mathbb{E}\left[X_{i,k}X_{i,l}\right] = \frac{1}{n}\left(\psi_{k,l}(0) + \sum_{d=1}^{m-1}\left(1 - \frac{d}{n}\right)(\psi_{l,k}(d) + \psi_{k,l}(d))\right), \qquad (3.2)$$

for $\psi_{k,l}(d) = \mathbb{P}\left[R_{d+1,k} \leq a_k, R_{1,l} \leq a_l\right] - \pi_k\pi_l$.

Let $X_i = (X_{i,1}, \cdots, X_{i,p})$ *for* $i = 1, \cdots, n$. *Note that* $\mathbb{A}_i = \{i-1, i, i+1\}$ *satisfies the requirement that* $j \notin \mathbb{A}_i \Rightarrow X_i, X_j$ *are independent and for all* $k, l \in \{1, \cdots, p\}$,

$$\mathbb{E}\left[X_{i,k}X_{i+1,l}\right] = \frac{1}{n^2}\sum_{d=1}^{m-1} d\psi_{k,l}(d). \tag{3.3}$$

Consider

$$\mathbf{Y}_n(t) = \sum_{i=1}^{\lfloor nt \rfloor}(X_{i,1}, \cdots, X_{i,p}) \quad t \in [0,1].$$

Let $\Sigma \in \mathbb{R}^{p\times p}$ *be given by*

$$\Sigma_{k,l} = \psi_{k,l}(0) + \sum_{d=1}^{m-1}\left(\psi_{l,k}(d) + \psi_{k,l}(d)\right). \tag{3.4}$$

We will bound the distance between $\mathbf{Y}_n$ *and* $\mathbf{Z} = \Sigma^{1/2}\mathbf{B}$, *where* $\mathbf{B}$ *is a standard p-dimensional Brownian Motion. Using the notation of Theorem 3.1, note that for all* $k \in \{1, \cdots, p\}$, $\lambda_k = n$, *for all* $i \in \{1, \cdots, n\}$, $J_{i,k} = \mathbb{1}_{[i/n,1]}$ *and*

(1) *By Cauchy-Schwarz and Jensen inequalities and (3.2),*

$$\begin{aligned}\epsilon_1 \leq & \frac{3}{2n^{1/2}}\sum_{k,l,r=1}^{p}\left\{\left(\psi_{k,k}(0) + 2\sum_{d=1}^{m-1}\left(1-\frac{d}{n}\right)(\psi_{k,k}(d))\right)^{1/2}\right.\\ & \cdot\left(\psi_{l,l}(0) + 2\sum_{d=1}^{m-1}\left(1-\frac{d}{n}\right)(\psi_{l,l}(d))\right)^{1/2}\\ & \left.\cdot\left(\psi_{r,r}(0) + 2\sum_{d=1}^{m-1}\left(1-\frac{d}{n}\right)(\psi_{r,r}(d))\right)^{1/2}\right\};\end{aligned}$$

(2) *By Cauchy-Schwarz and Jensen inequalities and (3.2),*

$$\begin{aligned}\epsilon_2 \leq & \frac{2}{3n^{1/2}}\sum_{k,l,r=1}^{p}\left\{\left(\psi_{k,k}(0) + 2\sum_{d=1}^{m-1}\left(1-\frac{d}{n}\right)(\psi_{k,k}(d))\right)^{1/2}\right.\\ & \cdot\left(\psi_{l,l}(0) + 2\sum_{d=1}^{m-1}\left(1-\frac{d}{n}\right)(\psi_{l,l}(d))\right)^{1/2}\\ & \left.\cdot\left(\psi_{r,r}(0) + 2\sum_{d=1}^{m-1}\left(1-\frac{d}{n}\right)(\psi_{r,r}(d))\right)^{1/2}\right\};\end{aligned}$$

(3) *By Cauchy-Schwarz and Jensen inequalities and (3.2),*

$$\epsilon_3 \leq \frac{2}{n^{1/2}} \sum_{k,l,r=1}^{p} \left\{ \left(\psi_{k,k}(0) + 2\sum_{d=1}^{m-1} \left(1-\frac{d}{n}\right)(\psi_{k,k}(d)) \right)^{1/2} \cdot \left(\psi_{l,l}(0) + 2\sum_{d=1}^{m-1} \left(1-\frac{d}{n}\right)(\psi_{l,l}(d)) \right)^{1/2} \cdot \left(\psi_{r,r}(0) + 2\sum_{d=1}^{m-1} \left(1-\frac{d}{n}\right)(\psi_{r,r}(d)) \right)^{1/2} \right\};$$

(4) *Since $K=1$ and by (3.2) and (3.4),*

$$\epsilon_4 = \frac{1}{2n} \sum_{k,l=1}^{p} \left| \sum_{d=1}^{m-1} d(\psi_{l,k}(d) + \psi_{k,l}(d)) \right|;$$

(5) *Since $K=1$ and by (3.3),*

$$\epsilon_5 \leq \frac{1}{n} \sum_{l,k=1}^{p} \sum_{d=1}^{m-1} d\psi_{k,l}(d);$$

(6) *By (3.4),*

$$\epsilon_6 = \frac{6\sqrt{5}p^{1/2}}{\sqrt{2\log 2}} \frac{\sqrt{\log(2n)}}{\sqrt{n}} \left[\sum_{k=1}^{p} \left(\psi_{k,k}(0) + 2\sum_{d=1}^{m-1} \psi_{k,k}(d) \right) \right]^{1/2};$$

(7) *Since for all $k \in \{1, \cdots, p\}$ and $i \in \{1, \cdots, n\}$, $J_{i,k} = \mathbb{1}_{[i/n,1]}$,*

$$\epsilon_7 = 0.$$

By Theorem 3.1, for any $g \in M^1$, as defined in (2.2),

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{Z})| \leq \|g\|_{M^1} \sum_{i=1}^{7} \epsilon_i,$$

which gives the desired bound. The bound clearly approaches zero as $n \to \infty$, which, by Proposition 2.3, implies that $\mathbf{Y}_n$ converges in distribution to $\mathbf{Z}$ with respect to the uniform topology.

The next result treats the case of independent p-dimensional terms with dependent components, whose scaled sum can be compared with a correlated p-dimensional Brownian Motion:

Proposition 3.5 (Dependent components). *Suppose that $X_1,...,X_n$, where $X_i=\left(X_i^{(1)},\cdots,X_i^{(p)}\right)$ for $i=1,\cdots,n$, are i.i.d. random vectors in $\mathbb{R}^p$. Suppose that each has a positive definite symmetric covariance matrix $\Sigma\in\mathbb{R}^{p\times p}$ and mean zero. Let:*

$$\mathbf{Y}_n(t)=n^{-1/2}\sum_{i=1}^{\lfloor nt\rfloor}X_i,\quad t\in[0,1]$$

and for $\mathbf{B}$, a standard p-dimensional Brownian Motion, let $\mathbf{Z}=\Sigma^{1/2}\mathbf{B}$. Then, for any $g\in M$:

$$\begin{aligned}&|\mathbb{E}g(\mathbf{Y}_n)-\mathbb{E}g(\mathbf{Z})|\\ \leq&\|g\|_M n^{-1/2}\left\{\sqrt{\log 2n}\left[\frac{6\sqrt{5}}{\sqrt{\pi}\log 2}\left(\sum_{i=1}^p\Sigma_{i,i}\right)^{1/2}+\frac{54\cdot 5^{1/3}p^{1/2}}{\sqrt{2}\log 2}\sum_{i=1}^p|\Sigma_{i,i}|^{3/2}\right]\right.\\ &+\frac{1}{6}\left(p^{1/2}\sum_{m=1}^p\mathbb{E}\left|X_1^{(m)}\right|^3+2\sum_{k,l=1}^p|\Sigma_{k,l}|\left(\sum_{m=1}^p\mathbb{E}\left|X_1^{(m)}\right|^2\right)^{1/2}\right)\\ &\left.+n^{-1}(\log 2n)^{3/2}p^{1/2}\frac{2160}{\sqrt{\pi}(\log 2)^{3/2}}\sum_{i=1}^p|\Sigma_{i,i}|^{3/2}\right\}.\end{aligned}$$

Remark 3.6. *If the components are uncorrelated and scaled in Proposition 3.5, i.e. $\Sigma=I_{p\times p}$, then the bound simplifies in the following way:*

$$\begin{aligned}&|\mathbb{E}g(\mathbf{Y}_n)-\mathbb{E}g(\mathbf{Z})|\\ \leq&\|g\|_M n^{-1/2}\left\{\sqrt{\log 2n}\left[\frac{6\sqrt{5}p^{1/2}}{\sqrt{2}\log 2}+\frac{54\cdot 5^{1/3}p^{3/2}}{\sqrt{\pi}\log 2}\right]\right.\\ &\left.+\frac{1}{6}\left(p^{1/2}\sum_{m=1}^p\mathbb{E}\left|X_1^{(m)}\right|^3+2p^{3/2}\right)+n^{-1}(\log 2n)^{3/2}p^{3/2}\frac{2160}{\sqrt{\pi}(\log 2)^{3/2}}\right\}.\end{aligned}$$

Remark 3.7. *For fixed p, by Proposition 2.3, Theorem 3.5 implies that $\mathbf{Y}_n\Rightarrow\mathbf{Z}$ in the uniform topology as the bound is of order $\frac{\sqrt{\log n}}{\sqrt{n}}$ in n. The bound also converges to 0 as $n\to\infty$ as long as $p=o\left(n^{1/5}\right)$.*

Remark 3.8 (Relevance of terms in the bound). *The first term appearing in the bound in Proposition 3.5 is an analogue of term ϵ_6 of Theorem 3.1. Similarly, the third and fourth term correspond to ϵ_1 and ϵ_3, respectively. The second and the last term are additional terms appearing due to the fact that M is larger than M^1 of (2.2). Since the summands have the limiting covariance structure a priori, no term corresponds to ϵ_4 and since the X_i's are multiplied by indicators $\mathbb{1}_{[i/n,1]}$ in the sum $\mathbf{Y}_n=n^{-1/2}X_i\mathbb{1}_{[i/n,1]}$, no term corresponds to ϵ_7 either. Due to independence between the summands, also no term corresponds to ϵ_2 and ϵ_5 (c.f. Remark 3.3).*

The next two result will be proved using ideas similar to those used to prove Theorem 3.1. It treats non-degenerate U-statistics. Those, as observed for instance in [Hal79, Corollary 1], after proper rescaling, represent a process created out of globally dependent summands and converge to standard Brownian Motion in distribution under certain conditions. We find a bound the rate of this convergence.

We note that in general U-statistics are defined to be random variables of the form:

$$S_n^k(h) = \sum_{1\leq i_1<...<i_k\leq n} h(X_{i_1},...,X_{i_k}), \quad n\geq 1$$

for a symmetric real (or complex) function h on $\mathcal{S}^k$ (where $\mathcal{S}$ is some measurable space) and a sequence of i.i.d. random variables $(X_i)_{i\geq 1}$ taking values in $\mathcal{S}$. Here, for simplicity, we only consider functions h on $\mathcal{S}^2$, yet our analysis can be readily extended to any $k\geq 2$. Also, we only consider non-degenerate U-statistics, i.e. those with $0<\sigma_w^2=\text{Var}(w(X_1))<\infty$, where $w(x)=\mathbb{E}[h(X_1,x)]$. The reason is that in the case of degenerate ones (i.e. those satisfying $\text{Var}(w(X_1))=0$) the limit in the invariance principle is non-Gaussian (see [Hal79, Corollary 1]), which is beyond the scope of this paper.

Theorem 3.9 (Non-degenerate bivariate U-statistics)**.** *Let $X_1,X_2,...$ be i.i.d. random variables taking values in some measurable space $\mathcal{S}$ and let $h:\mathcal{S}^2\to\mathbb{R}$ be a symmetric function such that $\mathbb{E}\left[h(X_1,X_2)\right]=0$, $\mathbb{E}\left[h^2(X_1,X_2)\right]=\sigma_h^2<\infty$. Also, suppose that, for the function $w(x)=\mathbb{E}[h(X_1,x)]$, we have that: $0<\sigma_w^2=Var(w(X_1))<\infty$ and $\mathbb{E}|w(X_1)|^3<\infty$. Let:*

$$\mathbf{Y}_n(t)=\frac{n^{-3/2}}{\sigma_w t}\sum_{1\leq i_1<i_2\leq\lfloor nt\rfloor} h(X_{i_1},X_{i_2}), \quad t\in[0,1]$$

and let $\mathbf{Z}$ be a standard Brownian Motion. Then, for any $g\in M^2$, as defined by (2.3):

$$|\mathbb{E}g(\mathbf{Y}_n)-\mathbb{E}g(\mathbf{Z})|\leq\|g\|_{M^2}n^{-1/2}\left[\left(141+16\frac{\sigma_h^2}{\sigma_w^2}+12\left(\frac{\sigma_h^2}{\sigma_w^2}-2\right)^{1/2}\right)\sqrt{\log 3n}\right.$$
$$\left.+43+\frac{\mathbb{E}|w(X_1)|^3+2\sigma_w^2\mathbb{E}|w(X_1)|}{6\sigma_w^3}\right].$$

Remark 3.10 (Relevance of terms in the bound)**.** *The term $\frac{\mathbb{E}|w(X_1)|^3+2\sigma_w^2\mathbb{E}|w(X_1)|}{6\sigma_w^3}$ appearing in the bound comes from the comparison of the process given by*

$$\tilde{\mathbf{Y}}_n(t)=\frac{n^{-3/2}}{\sigma_w t}\sum_{1\leq i_1<i_2\leq\lfloor nt\rfloor}\left(w(X_{i_1})+w(X_{i_2})\right), \quad t\in[0,1]$$

and a piecewise constant Gaussian process. It involves a Berry-Esseen-type third absolute moment component. The remaining terms come from the comparison of $\mathbf{Y}_n$ and $\tilde{\mathbf{Y}}_n$ and from the comparison of the piecewise constant Gaussian process and Brownian Motion for which the Brownian modulus of continuity is used.

Remark 3.11. *By Proposition 2.3, Theorem 3.9 implies that $\mathbf{Y}_n \Rightarrow \mathbf{Z}$ in the uniform (and Skorokhod) topology.*

Remark 3.12. *The bounds in Theorems 3.1, 3.9 and Proposition 3.5 are not optimised for constants as they are often estimated in a crude manner in the proofs presented in the section below. The constants are, however, expressed explicitly, which is often not the case in related pieces of literature.*

4. Proofs

The main tool used in the proofs of Theorems 3.1, 3.9 and Proposition 3.5 is Stein's method. It can be used in a surprisingly easy way to find a distance of the processes of interest from certain scaled sums of Gaussian random variables, which approximate the limiting continuous Gaussian process.

First, we set up Stein's method for distributions of certain D^p-valued random objects expressed as scaled sums of Gaussian random variables. Using a collection of Ornstein-Uhlenbeck processes with a Gaussian stationary law, we will construct a process whose stationary law is that of our target distribution. Then, we will find the infinitesimal generator $\mathcal{A}$ of that process and deduce that $\mathcal{A}g = g - \mathbb{E}_\mu g$ can be used as our Stein equation, where μ is the target law. This follows from the fact that $\mathbb{E}_\mu \mathcal{A}g = 0$ for all g in the domain of $\mathcal{A}$. We will then solve the Stein equation for all $g \in M$, using the analysis of [KDV17], and use some appealing properties of the Ornstein-Uhlenbeck semigroup to prove bounds on the derivatives of the solution.

4.1. Setting up Stein's method

Let $n, p \in \mathbb{N}_+$ and let $\tilde{Z}_{i,k}$'s be centred Gaussian random variables for $i = 1, \cdots, n$, $k = 1, \cdots, p$. Suppose that

a) the covariance matrix of $\left(\tilde{Z}_{1,1}, \cdots, \tilde{Z}_{1,p}, \tilde{Z}_{2,1}, \cdots, \tilde{Z}_{2,p}, \cdots, \tilde{Z}_{n,1}, \cdots, \tilde{Z}_{n,p}\right)$ is given by $\Sigma_n \in \mathbb{R}^{(np)\times(np)}$;
b) $J_{i,k} \in D([0,1], \mathbb{R})$, for $i = 1, \cdots, n$, $k = 1, \cdots, p$, are some functions independent of the $\tilde{Z}_{i,k}$'s.

Let

$$\mathbf{D}_n(t) = \left(\sum_{i=1}^{\lambda_1} \tilde{Z}_{i,1} J_{i,1}(t), \cdots, \sum_{i=1}^{\lambda_p} \tilde{Z}_{i,p} J_{i,p}(t)\right), \quad t \in [0,1], \tag{4.1}$$

Now let $\{(\mathscr{X}_{i,j}(u), u \geq 0) : i = 1, ..., n, j = 1, ...p\}$ be an array of i.i.d. Ornstein-Uhlenbeck processes with stationary law $\mathcal{N}(0,1)$, independent of the collection $\{J_{i,k} : i = 1, \cdots, n, k = 1, \cdots, p\}$. Consider:

$$\tilde{\mathscr{U}}(u) = (\Sigma_n)^{1/2} \left(\mathscr{X}_{1,1}(u), \cdots, \mathscr{X}_{1,p}(u), \mathscr{X}_{2,1}(u), \cdots, \mathscr{X}_{2,p}(u), \cdots, \mathscr{X}_{n,1}(u), \cdots \mathscr{X}_{n,p}(u)\right)^T$$

for $u \geq 0$ and write $\mathscr{U}_{i,j}(u) = \left(\tilde{\mathscr{U}}(u)\right)_{p(i-1)+j}$. Consider a process:

$$\mathbf{W}_n(t,u) = \left(\sum_{i=1}^{\lambda_1} \mathscr{U}_{i,1}(u) J_{i,1}(t), \cdots, \sum_{i=1}^{\lambda_p} \mathscr{U}_{i,p}(u) J_{i,p}(t)\right), \qquad t \in [0,1], \quad u \geq 0.$$

The stationary law of the process $(\mathbf{W}_n(\cdot,u))_{u\geq 0}$ is exactly the law of $\mathbf{D}_n$. We claim that:

Proposition 4.1. *The infinitesimal generator of the process $(\mathbf{W}_n(\cdot,u))_{u\geq 0}$ acts on any $f \in M$ in the following way:*

$$\mathcal{A}_n f(w) = -Df(w)[w] + \mathbb{E} D^2 f(w)\left[\mathbf{D}_n^{(2)}\right].$$

Remark 4.2. *The generator in Proposition 4.1 can also be written in the following way:*

$$\mathcal{A}_n f(w) = -Df(w)[w] + \sum_{k,l=1}^{p} \sum_{i=1}^{\lambda_k} \sum_{j=1}^{\lambda_l} (\Sigma_n)_{p(i-1)+k,p(j-1)+l} \mathbb{E} D^2 f(w)\left[e_k J_{i,k}, e_l J_{j,l}\right].$$

Let us prove a lemma that will be used in the proof of Proposition 4.1.

Lemma 4.3. *We have, for $u \geq 0, v \geq 0$:*

$$\mathbf{W}_n(\cdot,u+v) - e^{-v}\mathbf{W}_n(\cdot,u) \stackrel{\mathcal{D}}{=} \sigma(v)\mathbf{D}_n(\cdot)$$

for $\sigma^2(v) = 1 - e^{-2v}$.

Proof. We can construct i.i.d. standard Brownian Motions $\mathscr{B}_{i,j}$ such that $(\mathscr{X}_{i,j}(u), u \geq 0) = \left(e^{-u}\mathscr{B}_{i,j}(e^{2u}), u \geq 0\right)$. Then, writing $\mathbf{W}_n = \left(\mathbf{W}_n^{(1)}, \cdots, \mathbf{W}_n^{(p)}\right)$ and $\mathbf{D}_n = \left(\mathbf{D}_n^{(1)}, \cdots, \mathbf{D}_n^{(k)}\right)$ we obtain for all $k = 1, \cdots, p$:

$$\begin{aligned}
&\mathbf{W}_n^{(k)}(\cdot,u+v) - e^{-v}\mathbf{W}_n^{(k)}(\cdot,u)\\
=&\sum_{i=1}^{\lambda_k}\left[\mathscr{U}_{i,k}(u+v) - e^{-v}\mathscr{U}_{i,k}(u)\right] J_{i,k}(\cdot)\\
=&\sum_{i=1}^{\lambda_k}\left[\left(\tilde{\mathscr{U}}(u+v)\right)_{p(i-1)+k} - e^{-v}\left(\tilde{\mathscr{U}}(u)\right)_{p(i-1)+k}\right] J_{i,k}(\cdot)\\
=&\sum_{j=1}^{n}\sum_{l=1}^{p}\sum_{i=1}^{\lambda_k}\left(\Sigma_n^{1/2}\right)_{p(i-1)+k,p(j-1)+l}\left[\mathscr{X}_{j,l}(u+v) - e^{-v}\mathscr{X}_{j,l}(u)\right] J_{i,k}(\cdot)\\
\stackrel{\mathcal{D}}{=}&e^{-(u+v)}\sum_{j=1}^{n}\sum_{l=1}^{p}\sum_{i=1}^{\lambda_k}\left(\Sigma_n^{1/2}\right)_{p(i-1)+k,p(j-1)+l}\left[\mathscr{B}_{j,l}\left(e^{2(u+v)}\right) - \mathscr{B}_{j,l}\left(e^{2u}\right)\right] J_{i,k}(\cdot)\\
\stackrel{\mathcal{D}}{=}&\sigma(v)\mathbf{D}_n^{(k)}(\cdot),
\end{aligned}$$

as $\mathscr{B}_{j,l}\left(e^{2(u+v)}\right) - \mathscr{B}_{j,l}\left(e^{2u}\right) \sim \mathcal{N}\left(0, e^{2(u+v)} - e^{2u}\right)$. □

Proof of Proposition 4.1. Note that the semigroup of $(\mathbf{W}_n(\cdot,u))_{u\geq 0}$, acting on L is defined by:

$$(T_{n,u}f)(w) := \mathbb{E}\left[f\left(\mathbf{W}_n(\cdot,u)\right)\middle|\,\mathbf{W}_n(\cdot,0)=w\right] = \mathbb{E}\left[f(we^{-u}+\sigma(u)\mathbf{D}_n(\cdot)\right], \tag{4.2}$$

where the last equality follows from Lemma 4.3. By (4.2) and Lemma 2.1 we have that, for every $f\in M$:

$$\begin{aligned}&\Big|(T_{n,u}f)(w)-f(w)-\mathbb{E}Df(w)[\sigma(u)\mathbf{D}_n-w(1-e^{-u})]\\&-\frac{1}{2}\mathbb{E}D^2f(w)\left[\{\sigma(u)\mathbf{D}_n-w(1-e^{-u})\}^{(2)}\right]\Big|\\ \leq&\|f\|_M\mathbb{E}\|\sigma(u)\mathbf{D}_n-w(1-e^{-u})\|^3\\ \leq&K_1(1+\|w\|^3)u^{3/2}\end{aligned}$$

for a constant K_1 depending only on f, where the last inequality follows from the fact that for $u\geq 0$, $\sigma^3(u)\leq 3u^{3/2}$ and $(1-e^{-u})^3\leq u^{3/2}$. So:

$$\begin{aligned}&\left|(T_{n,u}f-f)(w)+uDf(w)[w]-u\mathbb{E}D^2f(w)[\mathbf{D}_n^{(2)}]\right|\\ \leq&\Big|(T_{n,u}f)(w)-f(w)-\mathbb{E}Df(w)[\sigma(u)\mathbf{D}_n-w(1-e^{-u})]\\&-\frac{1}{2}\mathbb{E}D^2f(w)[\{\sigma(u)\mathbf{D}_n-w(1-e^{-u})\}^{(2)}]\Big|+|\sigma(u)\mathbb{E}Df(w)[\mathbf{D}_n]|\\&+\left|(u-1+e^{-u})Df(w)[w]\right|+\left|\left(\frac{\sigma^2(u)}{2}-u\right)\mathbb{E}D^2f(w)[\mathbf{D}_n^{(2)}]\right|\\&+\left|\frac{(1-e^{-u})^2}{2}D^2f(w)[w^{(2)}]\right|+\left|\sigma(u)(1-e^{-u})\mathbb{E}D^2f(w)[\mathbf{D}_n,w]\right|\\ \leq&3u^{3/2}\left(K_2(1+\|w\|^3)+K_2(1+\|w\|^2)\|w\|+K_2(1+\|w\|)\mathbb{E}\|\mathbf{D}_n\|^2\right.\\&\left.+K_2(1+\|w\|)\|w\|^2+(1+\|w\|)\|w\|\mathbb{E}\|\mathbf{D}_n\|\right)+|\sigma(u)\mathbb{E}Df(w)[\mathbf{D}_n]|\\ \leq&K_3(1+\|w\|^3)u^{3/2},\end{aligned} \tag{4.3}$$

for some constants K_2 and K_3 depending only on f. The last inequality follows from the fact that:

$$\mathbb{E}Df(w)[\mathbf{D}_n]=\sum_{i=1}^p\sum_{k=1}^{\lambda_i}\left(s_n^{(i)}\right)^{-1}\mathbb{E}Df(w)\left[J_{k,i}e_i\right]\mathbb{E}[\tilde{Z}_{k,i}]=0.$$

Therefore, by (4.3), we obtain that:

$$\mathcal{A}_nf(w)=\lim_{u\searrow 0}\frac{T_{n,u}f(w)-f(w)}{u}=-Df(w)[w]+\mathbb{E}D^2f(w)\left[\mathbf{D}_n^{(2)}\right],$$

as required. □

Now we prove the following:

Proposition 4.4. *For any $g \in M$ such that $\mathbb{E}g(\mathbf{D}_n) = 0$, the Stein equation $\mathcal{A}_n f_n = g$ is solved by:*

$$f_n = \phi_n(g) = -\int_0^\infty T_{n,u} g du, \tag{4.4}$$

where $(T_{n,u}f)(w) = \mathbb{E}\left[f(we^{-u} + \sigma(u)\mathbf{D}_n(\cdot)\right]$. Furthermore:

$$\begin{aligned}
&A) \quad \|D\phi_n(g)(w)\| \leq \|g\|_M \left(1 + \frac{2}{3}\|w\|^2 + \frac{4}{3}\mathbb{E}\|\mathbf{D}_n\|^2\right),\\
&B) \quad \|D^2\phi_n(g)(w)\| \leq \|g\|_M \left(\frac{1}{2} + \frac{\|w\|}{3} + \frac{\mathbb{E}\|\mathbf{D}_n\|}{3}\right),\\
&C) \quad \frac{\left\|D^2\phi_n(g)(w+h) - D^2\phi_n(g)(w)\right\|}{\|h\|}\\
&\qquad \leq \sup_{w,h\in D^p} \frac{\|D^2(g+c)(w+h) - D^2(g+c)(w)\|}{3\|h\|}.
\end{aligned} \tag{4.5}$$

for any constant function $c : D^p \to \mathbb{R}$ and for all $w, h \in D^p$.

Remark 4.5. *It is worth noting that obtaining a small bound for $\mathbb{E}\|\mathbf{D}_n\|$ or $\mathbb{E}\|\mathbf{D}_n\|^2$ is not easy, unless $\mathbf{D}_n$ is a martingale and Doob's L^2 inequality can be used and $\mathbb{E}\|\mathbf{D}_n\|^2 \leq \mathbb{E}|\mathbf{D}_n(1)| = \mathbb{E}\sqrt{\sum_{i=1}^p \mathbf{D}_n^{(i)}(1)}$. This is, for instance, the case, if $\tilde{Z}_i = \left(\tilde{Z}_{i,1}, \cdots, \tilde{Z}_{i,p}\right)$'s are independent and $J_{i,k,n}$'s are independent.*

Proof. The first part of the proposition follows by the argument used to prove [KDV17, Proposition 1] upon noting that we can readily substitute $\mathbf{D}_n$ in the place of Z therein due to $\mathbb{E}\|\mathbf{D}_n\|^3$ being finite. This lets us conclude that the Stein equation $\mathcal{A}_n f_n = g$ is indeed solved by:

$$f_n = \phi_n(g) = -\int_0^\infty T_{n,u} g du.$$

Now, note that for ϕ_n defined in (4.4) we get:

$$\begin{aligned}
&\phi_n(g)(w+h) - \phi_n(g)(w)\\
&\overset{(4.2)}{=} -\mathbb{E}\int_0^\infty \left[g\left((w+h)e^{-u} + \sigma(u)\mathbf{D}_n\right) - g\left(we^{-u} + \sigma(u)\mathbf{D}_n\right)\right] du
\end{aligned}$$

and so dominated convergence (which can be applied because of [KDV17, (10)]) gives:

$$D^k\phi_n(g)(w) = -\mathbb{E}\int_0^\infty e^{-ku} D^k g(we^{-u} + \sigma(u)\mathbf{D}_n) du, \quad k = 1, 2. \tag{4.6}$$

Now, using (4.6) observe that:

A) $\|D\phi_n(g)(w)\|$

$$\begin{aligned}&\leq \int_0^\infty e^{-u}\mathbb{E}\|Dg(we^{-u}+\sigma(u)\mathbf{D}_n)\|du\\ &\leq \|g\|_M \int_0^\infty \left(e^{-u}+2\|w\|^2e^{-3u}+2\mathbb{E}\|\mathbf{D}_n\|^2(e^{-u}-e^{-3u})\right)du\\ &\leq \|g\|_M\left(1+\frac{2}{3}\|w\|^2+\frac{4}{3}\mathbb{E}\|\mathbf{D}_n\|^2\right),\end{aligned}$$

B) $\|D^2\phi_n(g)(w)\|$

$$\begin{aligned}&\leq \int_0^\infty e^{-2u}\mathbb{E}\left\|D^2g(we^{-u}+\sigma(u)\mathbf{D}_n)\right\|du\\ &\leq \|g\|_M \int_0^\infty e^{-2u}(1+\mathbb{E}\|we^{-u}+\sigma(u)\mathbf{D}_n\|)du\\ &\leq \|g\|_M\left(\frac{1}{2}+\frac{\|w\|}{3}+\frac{\mathbb{E}\|\mathbf{D}_n\|}{3}\right),\end{aligned}$$

C) $\dfrac{\|D^2\phi_n(g)(w+h)-D^2\phi_n(g)(w)\|}{\|h\|}$

$$\begin{aligned}&\leq \|h\|^{-1}\left\|\mathbb{E}\int_0^\infty e^{-2u}D^2g((w+h)e^{-u}+\sigma(u)\mathbf{D}_n)-e^{-2u}D^2g(we^{-u}+\sigma(u)\mathbf{D}_n)du\right\|\\ &\leq \sup_{w,h\in D^p}\frac{\|D^2g(w+h)-D^2g(w)\|}{\|h\|}\int_0^\infty e^{-2u}e^{-u}du\\ &=\sup_{w,h\in D^p}\frac{\|D^2(g+c)(w+h)-D^2(g+c)(w)\|}{3\|h\|},\end{aligned}$$

uniformly in $g\in M$, for any constant c, which proves (4.5). □

4.2. An auxiliary result

We now move to proving the main results of the paper. We start with an auxiliary lemma in which we use Stein's method combined with Taylor expansions to bound the distance between $\mathbf{Y}_n$, as defined in Theorem 3.1 and $\mathbf{D}_n$, as defined in (4.1). This result is of independent interest and will be used in all the proofs in this Section.

Lemma 4.6. *Consider the setup of Theorem 3.1. Let $\mathbf{D}_n$ be defined as in (4.1) for the covariance matrix Σ_n equal to the covariance matrix of $(X_{1,1},\cdots,X_{1,p},\cdots,X_{n,1},\cdots,X_{n,p})$. Let $g\in M$, as defined in Section 2. Then:*

$$|\mathbb{E}g(\mathbf{Y}_n)-\mathbb{E}g(\mathbf{D}_n)|$$

$$\leq \frac{\|g\|_M}{6}\sum_{i=1}^n\mathbb{E}\left\{\left(\sum_{k,l,m=1}^p\left[(X_{i,k})^2\|J_{i,k}\|^2\mathbb{1}_{[1,\lambda_k]}(i)\left(\sum_{j\in\mathbb{A}_i}X_{j,l}\|J_{j,l}\|\mathbb{1}_{[1,\lambda_l]}(j)\right)^2\right.\right.\right.$$

$$\left. \left. \left. \cdot \left(\sum_{j\in\mathbb{A}_i} X_{j,m} \|J_{j,m}\| \mathbb{1}_{[1,\lambda_m]}(j) \right)^2 \right] \right)^{1/2} \right\}$$

$$+ \frac{\|g\|_M}{3} \sum_{i=1}^n \sum_{j\in\mathbb{A}_i} \sum_{k,l=1}^p \mathbb{E}\left\{ \left[\sum_{m=1}^p \left(X_{i,k} \|J_{i,k}\| X_{j,l} \|J_{j,l}\| \mathbb{1}_{[1,\lambda_k]}(i) \mathbb{1}_{[1,\lambda_l]}(j) \right.\right.\right.$$

$$\left.\left.\left. \cdot \sum_{r\in\mathbb{A}_j\cap\mathbb{A}_i^c} X_{r,m} \|J_{r,m,n}\| \mathbb{1}_{[1,\lambda_m]}(r) \right)^2 \right]^{1/2} \right\}$$

$$+ \frac{\|g\|_M}{3} \sum_{i=1}^n \sum_{j\in\mathbb{A}_i} \sum_{k,l=1}^p \left\{ \mathbb{E}\left[|X_{i,k} X_{j,l}|\right] \mathbb{1}_{[1,\lambda_k]}(i) \mathbb{1}_{[1,\lambda_l]}(j) \right.$$

$$\left. \cdot \mathbb{E}\left[\|J_{i,k}\| \|J_{j,l}\| \sqrt{\sum_{m=1}^p \left(\sum_{r\in\mathbb{A}_i\cup\mathbb{A}_j} X_{r,m} \|J_{r,m}\| \mathbb{1}_{[1,\lambda_m]}(r) \right)^2} \right] \right\}.$$

The proof of Lemma 4.6 is based on manipulating the Stein operator, given in Proposition 4.4, using Taylor's theorem.

Proof of Lemma 4.6. Let $g_n = g - \mathbb{E}g(\mathbf{D}_n)$ and $f_n = \phi_n(g_n)$, as defined in (4.4). From Proposition 4.1 we know that:

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{D}_n)| = \left|\mathbb{E}\left[Df_n(\mathbf{Y}_n)\left[\mathbf{Y}_n\right] - D^2 f_n(\mathbf{Y}_n)\left[\mathbf{D}_n, \mathbf{D}_n\right]\right]\right|.$$

Let

$$\mathbf{Y}_n^j = \sum_{k\in\mathbb{A}_j^c} \left(X_{k,1}\mathbb{1}_{[1,\lambda_1]}(k) J_{k,1}, \cdots, X_{k,p}\mathbb{1}_{[1,\lambda_p]}(k) J_{k,p} \right)$$

and

$$\mathbf{Y}_n^{ij} = \sum_{k\in\mathbb{A}_j^c\cap\mathbb{A}_i^c} \left(X_{k,1}\mathbb{1}_{[1,\lambda_1]}(k) J_{k,1}, \cdots, X_{k,p}\mathbb{1}_{[1,\lambda_p]}(k) J_{k,p} \right).$$

Hence, $\mathbf{Y}_n^j$ is independent of X_j for all j and $\mathbf{Y}_n^{ij}$ is independent of X_i and X_j for all i, j. Therefore

$$\mathbb{E}Df_n(\mathbf{Y}_n^i)\left[\left(X_{i,1}\mathbb{1}_{[1,\lambda_1]}(i) J_{i,1}, , \cdots, X_{i,p}\mathbb{1}_{[1,\lambda_p]}(i) J_{i,p}\right)\right] = 0.$$

For $\{e_k : k = 1, \cdots, p\}$ denoting the elements of the canonical basis of $\mathbb{R}^p$ and for $i \in \{1, \cdots, n\}$,

$$
\begin{aligned}
&\Bigg|\mathbb{E}Df_n(\mathbf{Y}_n)\left[\left(X_{i,1}\mathbb{1}_{[1,\lambda_1]}(i)J_{i,1},\cdots,X_{i,p}\mathbb{1}_{[1,\lambda_p]}(i)J_{i,p}\right)\right]\\
&-\mathbb{E}\left[\sum_{j\in\mathbb{A}_i}\sum_{k,l=1}^{p}\left(X_{i,k}\mathbb{1}_{[1,\lambda_k]}(i)\right)\left(X_{j,l}\mathbb{1}_{[1,\lambda_l]}(j)\right)D^2f_n(\mathbf{Y}_n^i)\left[e_kJ_{i,k},e_lJ_{j,l}\right]\right]\Bigg|\\
=&\Bigg|\mathbb{E}Df_n(\mathbf{Y}_n)\left[\left(X_{i,1}\mathbb{1}_{[1,\lambda_1]}(i)J_{i,1},\cdots,X_{i,p}\mathbb{1}_{[1,\lambda_p]}(i)J_{i,p}\right)\right]\\
&-\mathbb{E}Df_n(\mathbf{Y}_n^i)\left[\left(X_{i,1}\mathbb{1}_{[1,\lambda_1]}(i)J_{i,1},\cdots,X_{i,p}\mathbb{1}_{[1,\lambda_p]}(i)J_{i,p}\right)\right]\\
&-\mathbb{E}D^2f_n(\mathbf{Y}_n^i)\Bigg[\left(X_{i,1}\mathbb{1}_{[1,\lambda_1]}(i)J_{i,1},\cdots,X_{i,p}\mathbb{1}_{[1,\lambda_p]}(i)J_{i,p}\right),\\
&\qquad\sum_{j\in\mathbb{A}_i}\left(X_{j,1}\mathbb{1}_{[1,\lambda_1]}(j)J_{j,1},\cdots X_{j,p}\mathbb{1}_{[1,\lambda_p]}(j)J_{j,p}\right)\Bigg]\Bigg|\\
\overset{(4.5)C)}{\leq}&\frac{\|g\|_M}{6}\mathbb{E}\Bigg[\left\|\left(X_{i,1}\mathbb{1}_{[1,\lambda_1]}(i)J_{i,1},\cdots,X_{i,p}\mathbb{1}_{[1,\lambda_p]}(i)J_{i,p}\right)\right\|\\
&\cdot\left\|\sum_{j\in\mathbb{A}_i}\left(X_{j,1}\mathbb{1}_{[1,\lambda_1]}(j)J_{j,1},\cdots,X_{j,p}\mathbb{1}_{[1,\lambda_p]}(j)J_{j,p}\right)\right\|^2\Bigg]\\
\leq&\frac{\|g\|_M}{6}\mathbb{E}\Bigg\{\Bigg(\sum_{k,l,m=1}^{p}\Bigg[\left(X_{i,k}\right)^2\|J_{i,k}\|^2\mathbb{1}_{[1,\lambda_k]}(i)\left(\sum_{j\in\mathbb{A}_i}X_{j,l}\|J_{j,l}\|\mathbb{1}_{[1,\lambda_l]}(j)\right)^2\\
&\cdot\left(\sum_{j\in\mathbb{A}_i}X_{j,m}\|J_{j,m}\|\mathbb{1}_{[1,\lambda_m]}(j)\right)^2\Bigg]\Bigg)^{1/2}\Bigg\}. \qquad (4.7)
\end{aligned}
$$

Furthermore, for all $i, j \in \{1, \cdots, n\}$,

$$
\begin{aligned}
&\big|\mathbb{E}\left[X_{i,k}\mathbb{1}_{[1,\lambda_k]}(i)X_{j,l}\mathbb{1}_{[1,\lambda_l]}(j)D^2f_n(\mathbf{Y}_n^i)\left[e_kJ_{i,k},e_lJ_{j,l}\right]\right]\\
&-\mathbb{E}\left[X_{i,k}\mathbb{1}_{[1,\lambda_k]}(i)X_{j,l}\mathbb{1}_{[1,\lambda_l]}(j)D^2f_n(\mathbf{Y}_n^{i,j})\left[e_kJ_{i,k},e_lJ_{j,l}\right]\right]\big|\\
\overset{(4.5)C)}{\leq}&\frac{\|g\|_M}{3}\mathbb{E}\Bigg\{\Bigg[\sum_{m=1}^{p}\Bigg(X_{i,k}\|J_{i,k}\|X_{j,l}\|J_{j,l}\|\mathbb{1}_{[1,\lambda_k]}(i)\mathbb{1}_{[1,\lambda_l]}(j)
\end{aligned}
$$

$$\left.\left.\left. \cdot \sum_{r\in\mathbb{A}_j\cap\mathbb{A}_i^c} X_{r,m}\|J_{r,m,n}\|\mathbb{1}_{[1,\lambda_m]}(r)\right)^2\right]^{1/2}\right\} \tag{4.8}$$

and

$$\begin{aligned}
&\left|\mathbb{E}\left[X_{i,k}\mathbb{1}_{[1,\lambda_k]}(i)X_{j,l}\mathbb{1}_{[1,\lambda_l]}(j)D^2f_n(\mathbf{Y}_n^{i,j})\left[e_kJ_{i,k},e_lJ_{j,l}\right]\right]\right.\\
&\left.-\mathbb{E}\left[X_{i,k}X_{j,l}\right]\mathbb{1}_{[1,\lambda_k]}(i)\mathbb{1}_{[1,\lambda_l]}(j)\mathbb{E}\left[D^2f_n(\mathbf{Y}_n)\left[e_kJ_{i,k},e_lJ_{j,l}\right]\right]\right|\\
\overset{(4.5)C)}{\leq}\;&\frac{\|g\|_M}{3}\mathbb{E}\left[|X_{i,k}X_{j,l}|\right]\mathbb{1}_{[1,\lambda_k]}(i)\mathbb{1}_{[1,\lambda_l]}(j)\\
&\cdot\mathbb{E}\left[\|J_{i,k}\|\,\|J_{j,l}\|\sqrt{\sum_{m=1}^p\left(\sum_{r\in\mathbb{A}_i\cup\mathbb{A}_j}X_{r,m}\|J_{r,m}\|\mathbb{1}_{[1,\lambda_m]}(r)\right)^2}\right].
\end{aligned}\tag{4.9}$$

Summing 4.7 over $i=1,\cdots,n$ and 4.8 and 4.9 over $i=1,\cdots,n$, $j\in\mathbb{A}_i$ and $k,l=1,\cdots,p$ will give us a bound on $|\mathbb{E}\mathcal{A}_ng(\mathbf{Y}_n)|$, as defined in Proposition 4.1, i.e. a bound on $|\mathbb{E}g(\mathbf{Y}_n)-\mathbb{E}g(\mathbf{D}_n)|$. □

4.3. Proof of Theorem 3.1

In the proof of Theorem 3.1 below, we will use the auxiliary processes $\mathbf{A}_n$ and $\tilde{\mathbf{A}}_n$, defined for $t\in[0,1]$ by

$$\mathbf{A}_n(t)=\left(\frac{1}{\sqrt{\lambda_1}}\sum_{i=1}^{\lambda_1}Z_{i,1}J_{i,1}(t),\cdots,\frac{1}{\sqrt{\lambda_p}}\sum_{i=1}^{\lambda_p}Z_{i,p}J_{i,p}(t)\right); \tag{4.10}$$

$$\tilde{\mathbf{A}}_n(t)=\left(\frac{1}{\sqrt{\lambda_1}}\sum_{i=1}^{\lambda_1}Z_{i,1}\mathbb{1}_{[i/\lambda_1,1]}(t),\cdots,\frac{1}{\sqrt{\lambda_p}}\sum_{i=1}^{\lambda_p}Z_{i,p}\mathbb{1}_{[i/\lambda_p,1]}(t)\right), \tag{4.11}$$

where $(Z_{i,1},\cdots,Z_{i,p})$'s, for $i=1,\cdots,n$, are i.i.d. Gaussian vectors with mean zero and covariance Σ, independent of the $J_{i,k}$'s and $X_{i,k}$'s for $i=1,\cdots,n$, $k=1,\cdots,p$.

In **Step 1** the distance between $\mathbf{D}_n$, as defined by (4.1), and $\mathbf{A}_n$ is bounded using bounds on the distance between two multivariate Gaussian distributions ([RR09, Proposition 2.8]). **Step 2** makes a straightforward use of the Mean Value Theorem to bound the distance between $\mathbf{A}_n$ and $\tilde{\mathbf{A}}_n$. In **Step 3** we couple $\tilde{\mathbf{A}}_n$ and $\mathbf{Z}$ in order to obtain a bound on $\mathbb{E}\|\tilde{\mathbf{A}}_n-\mathbf{Z}\|$ and then apply the Mean Value Theorem again to bound $|\mathbb{E}g(\tilde{\mathbf{A}}_n)-\mathbb{E}g(\mathbf{Z})|$ for all $g\in M^1$. Those three steps combined with Lemma 4.6 yield the assertion. In short:

$$|\mathbb{E}g(\mathbf{Y}_n)-\mathbb{E}g(\mathbf{Z})|\leq \underbrace{|\mathbb{E}g(\mathbf{Y}_n)-\mathbb{E}g(\mathbf{D}_n)|}_{\text{Lemma 4.6}}+\underbrace{|\mathbb{E}g(\mathbf{D}_n)-\mathbb{E}g(\mathbf{A}_n)|}_{\text{Step 1}}$$
$$+\underbrace{\left|\mathbb{E}g(\mathbf{A}_n)-\mathbb{E}g(\tilde{\mathbf{A}}_n)\right|}_{\text{Step 2}}+\underbrace{\left|\mathbb{E}g(\tilde{\mathbf{A}}_n)-\mathbb{E}g(\mathbf{Z})\right|}_{\text{Step 3}}.$$

Proof of theorem 3.1.
Step 1. Let $\lambda=\sum_{k=1}^p\lambda_k$ and consider function $f:\mathbb{R}^{\lambda_n}\to D^p[0,1]$ given by:

$$f\left(x_{1,1},\cdots,x_{\lambda_1,1},\cdots,x_{1,p},\cdots,x_{\lambda_p,p}\right)=\left(\sum_{i=1}^{\lambda_1}x_{i,1}J_{i,1,n},\cdots,\sum_{i=1}^{\lambda_p}x_{i,p}J_{i,p,n}\right).$$

This function is twice Fréchet differentiable with:

$$\begin{aligned}&\text{A)}\quad Df(x)[(h_{1,1},\cdots,h_{\lambda_1,1},\cdots,h_{1,p},\cdots,h_{\lambda_p,p})]\\&\quad=\left(\sum_{i=1}^{\lambda_1}h_{i,1}J_{i,1},\cdots,\sum_{i=1}^{\lambda_p}h_{i,p}J_{i,p}\right)\\&\text{B)}\quad D^2f(x)[h^{(1)},h^{(2)}]=0\end{aligned}$$

for all $x,h=(h_{1,1},\cdots,h_{\lambda_1,1},\cdots,h_{1,p},\cdots,h_{\lambda_p,p}),h^{(1)},h^{(2)}\in\mathbb{R}^{np}$. We notice that for the canonical basis vectors $e_i,e_j\in\mathbb{R}^{np}$ we have:

$$\left|D^2(g\circ f)(x)[e_i,e_j]\right|=\left|D^2g(f(x))[Df(x)[e_i],Df(x)[e_j]]\right|\leq K\sup_{w\in D}\|D^2g(w)\|$$

for all $x\in\mathbb{R}^{np}$ and K given in (3.1). Therefore, we can apply [RR09, Proposition 2.8] to the function $g\circ f$ and, recalling the definitions of $\mathbf{D}_n$ of (4.1) and $\mathbf{A}_n$ of (4.10), obtain

$$\begin{aligned}&|\mathbb{E}g(\mathbf{A}_n)-\mathbb{E}g(\mathbf{D}_n)|\\&\leq\frac{K}{2}\|g\|_{M^1}\sum_{k,l=1}^p\sum_{i=1}^{\lambda_k\wedge\lambda_l}\left[\sum_{j\in\mathbb{A}_i\setminus\{i\}}|\mathbb{E}\left[X_{i,k}X_{j,l}\right]|+\left|\frac{\Sigma_{k,l}}{\sqrt{\lambda_k\lambda_l}}-\mathbb{E}[X_{i,k}X_{i,l}]\right|\right],\end{aligned}\tag{4.12}$$

giving $\epsilon_4+\epsilon_5$.

Step 2. Also, note that, for $\mathbf{A}_n$ of (4.10) and $\tilde{\mathbf{A}}_n$ of (4.11),

$$\begin{aligned}
&\left|\mathbb{E}g(\mathbf{A}_n) - \mathbb{E}g(\tilde{\mathbf{A}}_n)\right| \\
\leq &\|g\|_{M^1}\mathbb{E}\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\| \\
\leq &\|g\|_{M^1}\mathbb{E}\left\{\sup_{t\in[0,1]}\sqrt{\sum_{k=1}^p\left[\frac{1}{\sqrt{\lambda_k}}\sum_{i=1}^{\lambda_k}Z_{i,k}\left(J_{i,k}(t) - \mathbb{1}_{[i/\lambda_k,1]}(t)\right)\right]^2}\right\} \\
\leq &\|g\|_{M^1}\sum_{k=1}^p\sum_{i=1}^{\lambda_k}\frac{1}{\sqrt{\lambda_k}}\mathbb{E}|Z_{i,k}|\mathbb{E}\left\|J_{i,k} - \mathbb{1}_{[i/\lambda_k,1]}\right\| \\
= &\|g\|_{M^1}\sum_{k=1}^p\sum_{i=1}^{\lambda_k}\frac{\sqrt{\Sigma_{i,i}}}{\sqrt{\lambda_k}}\mathbb{E}\left\|J_{i,k} - \mathbb{1}_{[i/\lambda_k,1]}\right\|, \qquad (4.13)
\end{aligned}$$

giving ϵ_7.

Step 3. We now realise a p-dimensional Brownian Motion $\mathbf{B}$ and let $\mathbf{Z} = \Sigma^{1/2}\mathbf{B}$. We also let

$$\tilde{\mathbf{A}}_n^{(j)}(t) = \mathbf{Z}^{(j)}(l/\lambda_j), \quad \text{for } t \in [l/\lambda_j, (l+1)/\lambda_j)$$

for every $j = 1, \cdots, p$, which agrees in distribution with our original definition (4.11) of $\tilde{\mathbf{A}}_n = \left(\tilde{\mathbf{A}}_n^{(1)}, \cdots, \tilde{\mathbf{A}}_n^{(p)}\right)$. Now, note that, using Jensen's inequality , we have:

$$\begin{aligned}
\mathbb{E}\|\tilde{\mathbf{A}}_n - \mathbf{Z}\| \leq &\left(\sum_{i=1}^p\mathbb{E}\left\|\tilde{\mathbf{A}}_n^{(i)} - \mathbf{Z}^{(i)}\right\|^2\right)^{1/2} \\
= &\sqrt{\sum_{i=1}^p\mathbb{E}\sup_{t\in[0,1]}\left|\mathbf{Z}^{(i)}(t) - \mathbf{Z}^{(i)}\left(\frac{\lfloor\lambda_i t\rfloor}{\lambda_i}\right)\right|^2} \\
\leq &\left(\sum_{i=1}^p\Sigma_{i,i}\right)^{1/2}\sqrt{\sum_{i=1}^p\mathbb{E}\sup_{t\in[0,1]}\left|\mathbf{B}^{(i)}(t) - \mathbf{B}^{(i)}\left(\frac{\lfloor\lambda_i t\rfloor}{\lambda_i}\right)\right|^2} \\
\leq &\frac{6\sqrt{5}}{\sqrt{2\log 2}}\left(\sqrt{\sum_{i=1}^p\frac{\log(2\lambda_i)}{\lambda_i}}\right)\left(\sum_{i=1}^p\Sigma_{i,i}\right)^{1/2},
\end{aligned}$$

where the third inequality follows because $\left\|\Sigma^{1/2}\right\|_2 = \sqrt{\lambda_{\max}(\Sigma)} \leq \left(\sum_{i=1}^p\Sigma_{i,i}\right)^{1/2}$, where $\lambda_{\max}(\Sigma)$ denotes the largest eigenvalue of Σ and the last inequality follows by [FN10, Lemma 3]. Therefore:

$$|\mathbb{E}g(\tilde{\mathbf{A}}_n) - \mathbb{E}g(\mathbf{Z})| \overset{\text{MVT}}{\leq} \sup_{w\in D^p}\|Dg(w)\|\mathbb{E}\|\mathbf{Z} - \tilde{\mathbf{A}}_n\|$$

$$\leq \|g\|_{M^1} \frac{6\sqrt{5}}{\sqrt{2\log 2}} \left(\sqrt{\sum_{i=1}^{p} \frac{\log(2\lambda_i)}{\lambda_i}} \right) \left(\sum_{i=1}^{p} \Sigma_{i,i} \right)^{1/2}, \tag{4.14}$$

giving ϵ_6.

Now, Lemma 4.6 (which gives $\epsilon_1 + \epsilon_2 + \epsilon_3$), combined with (4.12), (4.13), (4.14), yields the assertion. □

4.4. Proof of Proposition 3.5

The proof of Proposition 3.5 below is similar to that of Lemma 4.6 and **Step 3** of the proof of Theorem 3.1. Due to the independence of summands, the bound between $\mathbf{Y}_n$ and the pre-limiting Gaussian process has a simpler form than the one appearing in Theorem 3.1. We now work with all $g \in M$, contrary to what is done in the proof of Theorem 3.1. Hence, we need to bound both the first and second moment of the supremum distance between the pre-limiting process and the correlated Brownian Motion. This is necessary for the Mean Value Theorem to be applied in the final step.

Proof of Proposition 3.5. Let $\mathbf{D}_n$ be as in (4.1) with Σ_n such that the vectors $\left(\tilde{Z}_i\right)_{i=1}^n$ are independent and for all $i = 1, \cdots, n$ and $j = 1, \cdots, p$, $J_{i,j} = \mathbb{1}_{[i/n,1]}$. Let $g \in M$, $g_n = g - \mathbb{E}[g(\mathbf{D}_n)]$, $f_n = \phi_n(g_n)$, as in (4.4).

Note that for $\mathbf{Y}_n^j = \mathbf{Y}_n - \frac{1}{\sqrt{n}} X_j \mathbb{1}_{[j/n,1]}$, $\mathbf{Y}_n^j$ is independent of X_j and

$$\begin{aligned}
&\left| n^{-1/2}\mathbb{E}Df_n(\mathbf{Y}_n)\left[X_j\mathbb{1}_{[j/n,1]}\right] - n^{-1}\sum_{k,l=1}^{p}\Sigma_{k,l}\mathbb{E}D^2 f_n(\mathbf{Y}_n^j)\left[e_k\mathbb{1}_{[j/n,1]}, e_l\mathbb{1}_{[j/n,1]}\right]\right| \\
=&\Big| n^{-1/2}\mathbb{E}Df_n(\mathbf{Y}_n)\left[X_j\mathbb{1}_{[j/n,1]}\right] - n^{-1/2}\mathbb{E}Df_n(\mathbf{Y}_n^j)\left[X_j\mathbb{1}_{[j/n,1]}\right] \\
&-n^{-1}\mathbb{E}D^2 f_n(\mathbf{Y}_n^j)\left[X_j\mathbb{1}_{[j/n,1]}, X_j\mathbb{1}_{[j/n,1]}\right]\Big| \\
\leq& n^{-3/2}\frac{\|g\|_M}{6}\mathbb{E}\left\|X_j\mathbb{1}_{[j/n,1]}\right\|^3 \\
=& n^{-3/2}\frac{\|g\|_M}{6}\mathbb{E}\left[\left(\left(X_j^{(1)}\right)^2 + \cdots + \left(X_j^{(p)}\right)^2\right)^{3/2}\right] \\
\leq& p^{1/2}n^{-3/2}\frac{\|g\|_M}{6}\sum_{m=1}^{p}\mathbb{E}\left|X_j^{(m)}\right|^3,
\end{aligned} \tag{4.15}$$

where the first inequality follows by (4.5)C). Also, by (4.5)C):

$$\Bigg| n^{-1}\sum_{k,l=1}^{p}\Sigma_{k,l}\mathbb{E}D^2 f_n(\mathbf{Y}_n^j)\left[e_k\mathbb{1}_{[j/n,1]},e_l\mathbb{1}_{[j/n,1]}\right] - n^{-1}\sum_{k,l=1}^{p}\Sigma_{k,l}\mathbb{E}D^2 f_n(\mathbf{Y}_n)\left[e_k\mathbb{1}_{[j/n,1]},e_l\mathbb{1}_{[j/n,1]}\right]\Bigg| \leq n^{-3/2}\frac{\|g\|_M}{3}\sum_{k,l=1}^{p}|\Sigma_{k,l}|\left(\sum_{m=1}^{p}\mathbb{E}\left|X_j^{(m)}\right|^2\right)^{1/2}. \tag{4.16}$$

Let us now realise a p-dimensional Brownian Motion $\mathbf{B}$ and let $\mathbf{Z}=\Sigma^{1/2}\mathbf{B}$. Let us also define $\Sigma^{-1/2}\mathbf{D}_n(j/n)=\mathbf{B}(j/n)$ for every $j=1,...,n$, which agrees in distribution with our original definition of $\mathbf{D}_n$. Now, note that, by [FN10, Lemma 3] and Doob's L^3 inequality:

A) $$\mathbb{E}\|\mathbf{Z}-\mathbf{D}_n\|\leq\sqrt{\sum_{i=1}^{p}\mathbb{E}\left\|\mathbf{Z}^{(i)}-\mathbf{D}_n^{(i)}\right\|^2}\leq\frac{6\sqrt{5}}{\sqrt{2\log 2}}n^{-1/2}\sqrt{\log 2n}\left(\sum_{i=1}^{p}|\Sigma_{i,i}|\right)^{1/2};$$

B) $$\mathbb{E}\|\mathbf{Z}-\mathbf{D}_n\|^3\leq p^{1/2}\sum_{i=1}^{p}\mathbb{E}\|\mathbf{Z}^{(i)}-\mathbf{D}_n^{(i)}\|^3 \leq p^{1/2}\frac{1080}{\sqrt{\pi}(\log 2)^{3/2}}n^{-3/2}(\log 2n)^{3/2}\sum_{i=1}^{p}|\Sigma_{i,i}|^{3/2};$$

C) $$(\mathbb{E}\|\mathbf{Z}\|^3)^{2/3}\leq\left(p^{1/2}\sum_{i=1}^{p}\mathbb{E}\|\mathbf{Z}^{(i)}\|^3\right)^{2/3}\leq\frac{9p^{1/3}}{2\pi^{1/3}}\left(\sum_{i=1}^{p}|\Sigma_{i,i}|^{3/2}\right)^{2/3}.$$

Therefore:

$$\begin{aligned}
&|\mathbb{E}g(\mathbf{D}_n)-\mathbb{E}g(\mathbf{Z})|\\
\overset{\text{MVT}}{\leq}&\,\mathbb{E}\left[\sup_{c\in[0,1]}\|Dg(\tilde{\mathbf{Z}}+c(\mathbf{D}_n-\mathbf{Z}))\|\|\mathbf{Z}-\mathbf{D}_n\|\right]\\
\leq&\|g\|_M\mathbb{E}\left[\sup_{c\in[0,1]}(1+\|\mathbf{Z}+c(\mathbf{D}_n-\mathbf{Z})\|^2)\|\mathbf{Z}-\mathbf{D}_n\|\right]\\
\leq&\|g\|_M\left\{\mathbb{E}\|\mathbf{Z}-\mathbf{D}_n\|+2\mathbb{E}\|\mathbf{Z}-\mathbf{D}_n\|^3+2(\mathbb{E}\|\mathbf{Z}\|^3)^{2/3}(\mathbb{E}\|\mathbf{D}_n-\mathbf{Z}\|^3)^{1/3}\right\}\\
\leq&\|g\|_M\Bigg\{n^{-1/2}\sqrt{\log 2n}\left[\frac{6\sqrt{5}}{\sqrt{2\log 2}}\left(\sum_{i=1}^{p}|\Sigma_{i,i}|\right)^{1/2}+\frac{54\cdot 5^{1/3}p^{1/2}}{\sqrt{\pi}\log 2}\sum_{i=1}^{p}|\Sigma_{i,i}|^{3/2}\right]\\
&+n^{-3/2}(\log 2n)^{3/2}p^{1/2}\frac{2160}{\sqrt{\pi}(\log 2)^{3/2}}\sum_{i=1}^{p}|\Sigma_{i,i}|^{3/2}\Bigg\}.
\end{aligned}\tag{4.17}$$

We now sum (4.15) and (4.16) and sum them over j, which, combined with (4.17) yields the result. □

4.5. Proof of Theorem 3.9

In **Step 1** of the proof of Theorem 3.9 below, we consider a scaled sum of i.i.d random variables $w(X_i)$ and apply Lemma 4.6 together with an argument similar to **Step 2** and **Step 3** of the proof of Theorem 3.1 in order to bound the distance between this scaled sum and $\mathbf{Z}$. In **Step 2** we bound the distance between this scaled sum and our original process $\mathbf{Y}_n$ by bounding the second moment of the supremum distance between them and then using the Mean Value Theorem.

Proof of Theorem 3.9. Let $g \in M^2$.

Step 1. As in the proof of the invariance principle for U-statistics of [Hal79], we start by considering the behaviour of the following process $(\tilde{Y}_n(t), t \geq 0)$:

$$\tilde{\mathbf{Y}}_n(t) = \frac{n^{-3/2}}{\sigma_w t} \sum_{1 \leq i_1 < i_2 \leq \lfloor nt \rfloor} (w(X_{i_1}) + w(X_{i_2})) = \frac{1}{\sqrt{n}\sigma_w} \sum_{i=1}^{n} w(X_i) J_{i,n}(t),$$

where $J_{i,n}(t) = \frac{(\lfloor nt \rfloor - 1)\mathbb{1}_{[i/n,1]}(t)}{nt}$. Recall that $w(x) = \mathbb{E}h(X_1, x)$. Let $\mathbf{A}_n(t) = n^{-1/2}\sum_{i=1}^n Z_i J_{i,n}(t)$ and $\tilde{\mathbf{A}}_n(t) = n^{-1/2}\sum_{i=1}^{\lfloor nt \rfloor} Z_i$, where $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$.

Note that Lemma 4.6 readily yields that:

$$\left|\mathbb{E}g(\hat{\mathbf{Y}}_n) - \mathbb{E}g(\mathbf{A}_n)\right| \leq \frac{\|g\|_M}{6\sigma_w^3} n^{-1/2} \left(\mathbb{E}\left|w(X_1)\right|^3 + 2\sigma_w^2 \mathbb{E}|w(X_1)|\right), \tag{4.18}$$

as $\|J_{i,n}\| \leq 1$ for all $i, n \in \mathbb{N}$ and $w(X_i)$'s for $i = 1, \cdots, n$ are independent.

We see that, by Doob's L^2 inequality, we have for every m:

$$\mathbb{E}\left[\max_{1 \leq l \leq m}\left|\sum_{i=1}^{l} Z_i\right|\right]^2 \leq 4m = 4\sum_{i=1}^{m} 1.$$

Therefore, using [Faz14, Theorem1] for inequality $(*)$, we obtain:

$$\begin{aligned}
\text{A)} \quad \mathbb{E}\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\|^2 \leq & n^{-1}\mathbb{E}\left[\max_{1 \leq l \leq n}\left|\frac{l-1}{l+1} - 1\right|\left|\sum_{i=1}^{l} Z_i\right|\right]^2 \\
\leq & n^{-1} 2\mathbb{E}\left[\max_{1 \leq l \leq n}\left|\frac{\sum_{i=1}^{l} Z_i}{l+1}\right|\right]^2 \\
\overset{(*)}{\leq} & 32 n^{-1} \sum_{i=1}^{n} \frac{1}{i^2} \\
\leq & \frac{16\pi^2}{3} n^{-1} \\
\text{B)} \quad \mathbb{E}\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\| \leq & \sqrt{\mathbb{E}\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\|^2} \leq \frac{4\pi}{\sqrt{3}} n^{-1/2}.
\end{aligned} \tag{4.19}$$

Doob's L^2 inequality readily gives us:

$$\mathbb{E}\|\tilde{\mathbf{A}}_n\|^2 = \mathbb{E}\left[\left(\max_{1\leq m\leq n}\frac{m-1}{m}n^{-1}\left|\sum_{i=1}^m Z_i\right|\right)^2\right] \leq 4. \tag{4.20}$$

It follows that:

$$\begin{aligned}
&|\mathbb{E}g(\mathbf{A}_n) - \mathbb{E}g(\tilde{\mathbf{A}}_n)| \\
\leq &\mathbb{E}\left[\sup_{c\in[0,1]}\|Dg((1-c)\tilde{\mathbf{A}}_n + c\mathbf{A}_n)\|\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\|\right] \\
\leq &\|g\|_{M^2}\mathbb{E}\left[\sup_{c\in[0,1]}(1+\|\tilde{\mathbf{A}}_n + c(\mathbf{A}_n - \tilde{\mathbf{A}}_n)\|)\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\|\right] \\
\leq &\|g\|_{M^2}\left(\mathbb{E}\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\| + \mathbb{E}\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\|^2 + \sqrt{\mathbb{E}\|\tilde{\mathbf{A}}_n\|^2}\sqrt{\mathbb{E}\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\|^2}\right) \\
\leq &\|g\|_{M^2}\left(\frac{12\pi}{\sqrt{3}}n^{-1/2} + \frac{16\pi^2}{3}n^{-1}\right),
\end{aligned} \tag{4.21}$$

where the first inequality follows from the Mean Value Theorem and the last one follows from (4.19) and (4.20). Also, by [FN10, Lemma 3] and Doob's L^2 inequality:

A) $\mathbb{E}\|\mathbf{A}_n - \mathbf{Z}\| \leq \frac{30}{\sqrt{\pi\log 2}}n^{-1/2}\sqrt{\log 2n}$

B) $\mathbb{E}\|\mathbf{A}_n - \mathbf{Z}\|^2 \leq \frac{90}{\log 2}n^{-1}\log 2n$

C) $\mathbb{E}\|\mathbf{Z}\|^2 \leq 4$

and therefore:

$$\begin{aligned}
&|\mathbb{E}g(\mathbf{A}_n) - \mathbb{E}g(\mathbf{Z})| \\
\leq &\|g\|_{M^2}\left(\mathbb{E}\|\mathbf{A}_n - \mathbf{Z}\| + \mathbb{E}\|\mathbf{A}_n - \mathbf{Z}\|^2 + \sqrt{\mathbb{E}\|\mathbf{Z}\|^2}\sqrt{\mathbb{E}\|\mathbf{A}_n - \mathbf{Z}\|^2}\right) \\
\leq &\|g\|_{M^2}n^{-1/2}\left[\left(\frac{30}{\sqrt{\pi\log 2}} + \frac{12\sqrt{5}}{\sqrt{2\log 2}}\right)\sqrt{\log 2n} + \frac{90}{\log 2}n^{-1/2}\log 2n\right].
\end{aligned} \tag{4.22}$$

Step 2. We now wish to find a bound on $|\mathbb{E}g(\tilde{\mathbf{Y}}_n) - \mathbb{E}g(\mathbf{Y}_n)|$. Note that:

$$\mathbf{Y}_n - \tilde{\mathbf{Y}}_n = \frac{n^{-3/2}}{\sigma_w t}\sum_{1\leq i_1 < i_2 \leq \lfloor nt\rfloor}\left(h(X_{i_1}, X_{i_2}) - w(X_{i_1}) - w(X_{i_2})\right).$$

Let $\phi_h^2 = \mathbb{E}h^2(X_1, X_2)$. First, note that, if $\mu = \mathcal{L}(X_1)$ (i.e. μ is the law of X_1),

$$\begin{aligned}
&\mathbb{E}\left[(h(X_1,X_2)-w(X_1)-w(X_2))\left(h(X_1,X_3)-w(X_1)-w(X_3)\right)\right]\\
=&\mathbb{E}\left[h(X_1,X_2)h(X_1,X_3)\right]-2\mathbb{E}\left[h(X_1,X_2)w(X_1)\right]+\mathbb{E}w^2(X_1)\\
=&\int\int\int h(x,y)h(x,z)\mu(dx)\mu(dy)\mu(dz)\\
&-2\int\int h(x,y)\int h(x,z)\mu(dz)\mu(dx)\mu(dy)\\
&+\int\int h(x,y)\mu(dy)\int h(x,z)\mu(dz)\mu(dx)=0,
\end{aligned}$$

where the first equality follows by the fact that $w(X_2)$ is independent of $h(X_1, X_3)$, $w(X_1)$ and $w(X_3)$, $w(X_3)$ is independent of $h(X_1, X_2)$, $w(X_1)$ and $w(X_2)$, and $\mathbb{E}w(X_2) = \mathbb{E}w(X_3) = 0$. Therefore:

$$\begin{aligned}
&\mathbb{E}\left[\sum_{1\leq i_1<i_2\leq m}\left(h(X_{i_1},X_{i_2})-w(X_{i_1})-w(X_{i_2})\right)\right]^2\\
=&\binom{m}{2}\mathbb{E}\left[h(X_1,X_2)-w(X_1)-w(X_2)\right]^2\\
=&\binom{m}{2}\left[\sigma_h^2+2\sigma_w^2-4\int\int h(x,y)\int h(x,z)\mu(dz)\mu(dx)\mu(dy)\right]\\
=&\binom{m}{2}(\sigma_h^2-2\sigma_w^2). \qquad (4.23)
\end{aligned}$$

Now, $\sum_{1\leq i_1<i_2\leq m}\left(h(X_{i_1},X_{i_2})-w(X_{i_1})-w(X_{i_2})\right)$ is a martingale with respect to the filtration $\sigma(X_1, ..., X_m)$. Indeed:

$$\begin{aligned}
&\mathbb{E}\left[\left.\sum_{1\leq i_1<i_2\leq m+1}\left(h(X_{i_1},X_{i_2})-w(X_{i_1})-w(X_{i_2})\right)\right| X_1,...,X_m\right]\\
=&\sum_{1\leq i_1<i_2\leq m}\left(h(X_{i_1},X_{i_2})-w(X_{i_1})-w(X_{i_2})\right)\\
&+\mathbb{E}\left[\left.\sum_{i=1}^m\left(h(X_i,X_{m+1})-w(X_i)-w(X_{m+1})\right)\right| X_1,...,X_m\right]\\
=&\sum_{1\leq i_1<i_2\leq m}\left(h(X_{i_1},X_{i_2})-w(X_{i_1})-w(X_{i_2})\right)+\sum_{i=1}^m\left(\mathbb{E}\left[\left.h(X_i,X_{m+1})\right| X_i\right]-w(X_i)\right)\\
=&\sum_{1\leq i_1<i_2\leq m}\left(h(X_{i_1},X_{i_2})-w(X_{i_1})-w(X_{i_2})\right).
\end{aligned}$$

Hence, Doob's inequalities give us, for every m, such that $1 \leq m \leq n$:

$$\mathbb{E}\left[\max_{1\leq l\leq m}\left|\sum_{1\leq i_1<i_2\leq l}\left(h(X_{i_1},X_{i_2})-w(X_{i_1})-w(X_{i_2})\right)\right|\right]^2\overset{(4.23)}{\leq}4\binom{m}{2}(\sigma_h^2-2\sigma_w^2).$$

Then, by [Faz14, Theorem 1], applied with $\beta_i = \alpha_i = i$ and $r = 2$, and using the fact that $\binom{m}{2} = \sum_{i=1}^{m}(i-1)$, we obtain:

$$\begin{aligned}\mathbb{E}\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\|^2 =& \frac{n^{-3}}{\sigma_w^2}\mathbb{E}\left[\sup_{t\in[0,1]}\left|t^{-1}\sum_{1\leq i_1<i_2\leq\lfloor nt\rfloor}(h(X_{i_1},X_{i_2}) - w(X_{i_1}) - w(X_{i_2}))\right|^2\right]\\ =& \frac{n^{-1}}{\sigma_w^2}\mathbb{E}\left[\max_{1\leq l\leq n} l^{-1}\left|\sum_{1\leq i_1<i_2\leq l}(h(X_{i_1},X_{i_2}) - w(X_{i_1}) - w(X_{i_2}))\right|\right]^2\\ \leq& 16\left(\frac{\sigma_h^2}{\sigma_w^2} - 2\right)\sum_{i=1}^{n}\frac{1}{i}n^{-1} \leq 16\left(\frac{\sigma_h^2}{\sigma_w^2} - 2\right)n^{-1}\log 3n. \end{aligned}\tag{4.24}$$

Also, by Doob's L^2 inequality:

$$\mathbb{E}\|\tilde{\mathbf{Y}}_n\|^2 = n^{-3}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor - 1}{t}\sum_{i=1}^{\lfloor nt\rfloor}\frac{w(X_i)}{\sigma_w}\right|\right]^2 = n^{-1}\mathbb{E}\left[\sup_{1\leq l\leq n}\left|\frac{l-1}{l}\sum_{i=1}^{l}\frac{w(X_i)}{\sigma_w}\right|\right]^2 \leq 4. \tag{4.25}$$

Therefore:

$$\begin{aligned}&|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\tilde{\mathbf{Y}}_n)|\\ \leq& \mathbb{E}\left[\sup_{c\in[0,1]}\|Dg\left((1-c)\tilde{\mathbf{Y}}_n + c\mathbf{Y}_n\right)\|\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\|\right]\\ \leq& \|g\|_{M^2}\mathbb{E}\left[\sup_{c\in[0,1]}(1 + \|\tilde{\mathbf{Y}}_n + c(\mathbf{Y}_n - \tilde{\mathbf{A}}_n)\|)\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\|\right]\\ \leq& \|g\|_{M^2}\left(\mathbb{E}\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\| + \mathbb{E}\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\|^2 + \sqrt{\mathbb{E}\|\tilde{\mathbf{Y}}_n\|^2}\sqrt{\mathbb{E}\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\|^2}\right)\\ \leq& \|g\|_{M^2}\left(12\left(\frac{\sigma_h^2}{\sigma_w^2} - 2\right)^{1/2} n^{-1/2}\sqrt{\log 3n} + 16\left(\frac{\sigma_h^2}{\sigma_w^2} - 2\right)n^{-1}\log 3n\right), \end{aligned}\tag{4.26}$$

where the first inequality follows from the Mean Value Theorem and the last one follows by (4.24) and (4.25).

We combine (4.18), (4.21), (4.22) and (4.26) to obtain the assertion. □

Remark 4.7. *While, in the proof of Theorem 3.9 above, it is possible to obtain a bound on $|\mathbb{E}g(\tilde{\mathbf{Y}}_n) - \mathbb{E}g(\mathbf{Z})|$ for any $g \in M$, using methods analogous to those which let us prove Theorem 3.1, the situation becomes more complicated when it comes to approximating the remainder. This is because using Doob's L^3 inequality and [Faz14, Corollary 1] for $\mathbb{E}\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\|^3$ gives a bound which does not converge to 0 with n. Therefore, in (4.26) we cannot go beyond the second moment of $\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\|$. Hence, for our technique of proof, it is necessary that we assume $g \in M^2$, as defined by (2.3).*

Remark 4.8. *The stronger assumption of $g \in M^1$ in Theorem 3.9 would simplify its proof. Namely, using the notation of the proof of Theorem 3.9, we could treat $\hat{Y}_n$ as a scaled sum of i.i.d. mean zero, variance 1 random variables $\frac{w(X_i)}{\sigma_w}$. Using (4.19) and applying Theorem 3.1 gives:*

$$\left|\mathbb{E}g(\tilde{\mathbf{Y}}_n) - \mathbb{E}g(\mathbf{Z})\right| \leq \frac{\|g\|_{M^1}}{2} n^{-1/2} \left(\frac{\mathbb{E}|w(X_1)|^3}{\sigma_w^3} + 8 + 10\sqrt{\log 2n}\right)$$

and (4.26) could be substituted with:

$$\left|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\tilde{\mathbf{Y}}_n)\right| \leq \|g\|_{M^1}\mathbb{E}\|\mathbf{Y}_n - \tilde{\mathbf{Y}}_n\| \overset{(4.24)}{\leq} \|g\|_M \left(\frac{\sigma_h^2}{\sigma_w^2} - 2\right)^{1/2} \frac{4\sqrt{\log 3n}}{n^{1/2}}.$$

Acknowledgements

The author would like to thank Gesine Reinert for many helpful discussions and constructive comments on the early versions of this paper. The author is also grateful to Giovanni Peccati and Christian Döbler for spotting a mistake in the proof of Lemma 4.9 and suggesting an alternative approach to proving it.

The author has been supported by the EPSRC PhD studentship (reference number 1654155).

References

[Bar90] A.D. Barbour. Stein's Method for Diffusion Approximation. *Probability Theory and Related Fields*, 84:297–322, 1990.

[BDM18] E. Besançon, L. Decreusefond, and P. Moyal. Stein's method for diffusive limit of Markov processes. arXiv:1805.01691, 2018.

[BHJ92] A.D. Barbour, L. Holst, and S. Janson. *Poisson Approximation*. Oxford Studies in Probability. Clarendon Press, 1992.

[Bil99] P. Billingsley. *Convergence of Probability Measures, 2nd Edition*. Wiley Series in Probability and Statistics. Wiley-Blackwell, 1999.

[BJ09] A.D. Barbour and S. Janson. A functional combinatorial central limit theorem. *Electronic Journal of Probability*, 14(81):2352–2370, 2009.

[CD13] L. Coutin and L. Decreusefond. Stein's method for Brownian Approximations. *Communications on Stochastic Analysis*, 7(3):349–372, 2013.

[CGS11] L.H.Y Chen, L. Goldstein, and Q.-M. Shao. *Normal Approximation by Stein's Method*. Probability and Its Applications. Springer Verlag, 2011.

[Chr87] T.C. Christofides. *Maximal probability inequalities for multidimensionally indexed semimartingales and convergence theory of u-statistics*. PhD thesis, Johns Hopkins University, 1987.

[CS07] L.H.Y. Chen and Q.-M. Shao. Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli*, 13(2):581–599, 05 2007.

[dJ90] P. de Jong. A central limit theorem for generalized multilinear forms. *Journal of Multivariate Analysis*, 34(2):275 – 289, 1990.

[DK92] A. Dembo and S. Karlin. Poisson approximations for r-scan processes. *Ann. Appl. Probab.*, 2(2):329–357, 05 1992.

[Don51] M.D. Donsker. An invariance principle for certain probability limit theorems. *Memoirs of the American Mathematical Society*, 6, 1951.

[DP17] Ch. Döbler and G. Peccati. Quantitative de jong theorems in any dimension. *Electron. J. Probab.*, 22:35 pp., 2017.

[DR96] A. Dembo and Y. Rinott. Some examples of normal approximations by stein's method. In D. Aldous and R. Pemantle, editors, *Random Discrete Structures*, pages 25–44, New York, NY, 1996. Springer New York.

[EL99] P. Eichelsbacher and M. Lowe. Large deviations in partial sums of U-processes. *Theory Probab. Appl.*, 43(1):26–41, 1999.

[Faz14] I. Fazekas. On a general approach to the Strong Law of Large Numbers. *Journal of Mathematical Sciences*, 200(4):411–423, 2014.

[FN10] M. Fischer and G. Nappo. On the Moments of the Modulus of Continuity of Ito Processes. *Stochastic Analysis and Applications*, 28(1):103–122, 2010.

[GNW01] J. Glaz, J. Naus, and S. Wallenstein. *Scan Statistics*. Springer Series in Statistics. Springer-Verlag New York, 2001.

[Hal79] P. Hall. On the invariance principle for U-statistics. *Stochastic Processes and Their Applications*, 9(2):163–174, 1979.

[Hoe48] W. Hoeffding. A Class of Statistics with Asymptotically Normal Distribution. *Ann. Math. Statist.*, 19(3):293–325, 1948.

[Hoe61] W. Hoeffding. *The strong law of large numbers for U-statistics*. Institute of Statistics mimeo series 302. North Carolina State University. Dept. of Statistics, 1961.

[Jan97] S. Janson. *Gaussian Hilbert Spaces*. Cambridge Tracts in Mathematics. Cambridge University Press, 1997.

[Kas17a] M.J. Kasprzak. Diffusion approximations via Stein's method and time changes. arXiv:1701.07633, 2017.

[Kas17b] M.J. Kasprzak. Multivariate functional approximations with Stein's method of exchangeable pairs. arXiv:1710.09263, 2017.

[KB92] S. Karlin and V. Brendel. Chance and statistical significance in protein and dna sequence analysis. *Science*, 257(5066):39–49, 1992.

[KDV17] M.J. Kasprzak, A. B. Duncan, and S.J. Vollmer. Note on A. Barbour's paper on Stein's method for diffusion approximations. *Electron. Commun. Probab.*, 22(23):1–8, 2017.

[KJ88] S. Kotz and N.L. Johnson, editors. *U-statistics*, volume 9 of *Encyclopedia of Statistical Sciences*, pages 436–444. John Wiley and Sons, Inc., 1988.

[LRS17] C. Ley, G. Reinert, and Y. Swan. Stein's method for comparison of univariate distributions. *Probability Surveys*, 14:1–52, 2017.

[Nau82] J. Naus. Approximations for Distributions of Scan Statistics. *Journal of the American Statistical Association*, 77(377):177–183, 1982.

[NP12] I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus*. Cambridge tracts in Mathematics. Cambridge University Press, 2012.

[Ros11] N. Ross. Fundamentals of Stein's Method. *Probability Surveys*, 8:210–293, 2011.

[RR97] Y. Rinott and V. Rotar. On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.*, 7(4):1080–1105, 11 1997.

[RR09] G. Reinert and A. Röllin. Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *The Annals of Probability*, 37(6):2150–2173, 2009.

[RV80] H. Rubin and R.A. Vitale. Asymptotic Distribution of Symmetric Statistics. *Ann. Statist.*, 8(1):165–170, 1980.

[Ser80] R.J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. John Wiley and Sons, Inc., 1980.

[Shi11] H.-H. Shih. On Stein's method for infinite-dimensional Gaussian approximation in abstract Wiener spaces. *Journal of Functional Analysis*, 261(5):1236 – 1283, 2011.

[Ste72] Ch. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, 2:583–602, 1972.

[Swa16] Y. Swan. A gateway to Stein's Method. https://sites.google.com/site/steinsmethod/home, 2016. Accessed on 19/05/2016.

[Vit84] A.R. Vitale. *An expansion for symmetric statistics and the Efron-Stein inequality*, volume 5 of *Lecture Notes–Monograph Series*, pages 112–114. Institute of Mathematical Statistics, 1984.

[WW40] A. Wald and J. Wolfowitz. On a test whether two samples are from the same population. *Ann. Math. Statist.*, 11(2):147–162, 06 1940.

Appendix. Proof of Proposition 2.3

As in the proof of [BJ09, Proposition 3.1], we note that, by Skorokhod's representation theorem, $\mathbf{Z}_n$ and $\mathbf{Z}$ can be defined on the same probability space in such a way that $\|\mathbf{Z}_n - \mathbf{Z}\| \xrightarrow{n\to\infty} 0$ a.s. (as $\mathbf{Z}$ is continuous). The fact that $C([0,1],\mathbb{R}^p)$ equipped with norm $\|\cdot\|$ is separable, by the Stone-Weierstrass theorem, lets us use the argument of the proof of the Skorokhod representation theorem presented in [Bil99, Chapter 5] and conclude that it is enough to show that $\mathbb{P}[\mathbf{Y}_n \in B] \to \mathbb{P}[\mathbf{Z} \in B]$ for all sets $B = \bigcap_{1\leq l\leq L} B_l$, where $B_l = \{w \in D^p : \|w - s_l\| < \gamma_l\}$, $s_l \in C([0,1],\mathbb{R}^p)$ and γ_l is such that $\mathbb{P}[\mathbf{Z} \in \partial B_l] = 0$. Let us fix such a set B.

Let $\phi : \mathbb{R}^+ \to [0,1]$ be a non-increasing, three times continuously differentiable function satisfying, $\phi(x) = 1$ for $x \leq 0$ and $\phi(x) = 0$ for $x \geq 1$ and fix some $0 < \epsilon, \eta_n \leq 1, p_n \geq 4$. Define $g_{l,n} : D^p \to \mathbb{R}$ by:

$$g_{l,n}(w) = \phi\left(\frac{\left\|\sqrt{(\epsilon\gamma_l)^2 + \sum_{i=1}^p \left((w - s_l)^{(i)}\right)^2}\right\|_{p_n} - \gamma_l\sqrt{1+\epsilon^2}}{\eta_n}\right),$$

where $\|w\|_{p_n} := \left(\int_0^1 |w(t)|^{p_n} dt\right)^{1/p_n}$ for any $w \in D^p$. We have the following result:

Lemma 4.9. *For any finite L:*

$$\left\|\prod_{l=1}^L g_{l,n}\right\|_{M^0} \leq \tilde{C} p_n^2 \eta_n^{-3}. \tag{4.27}$$

for a constant $\tilde{C}$ independent of p_n and η_n (which might depend on ϵ or γ_l's).

Proof. First, ϕ, ϕ', ϕ'', ϕ''' are all everywhere continuous and constant outside of the compact interval $[0,1]$ and therefore bounded. Therefore also $\frac{|\phi''(x+h)-\phi''(x)|}{|h|}$ must be uniformly bounded.

Furthermore, let

$$f(w) = \frac{\left\|\sqrt{(\epsilon\gamma_l)^2 + \sum_{i=1}^p \left((w - s_l)^{(i)}\right)^2}\right\|_{p_n}}{\eta_n}, \tag{4.28}$$

and denote by $|\cdot|$ the Euclidean norm, and by $\langle\cdot\rangle$ the Euclidean inner product.

Step 1: Bounding the first derivative of f of (4.28)

We have that

$$\begin{aligned} Df(w)[h] =& \frac{1}{p_n\eta_n}\left(\int_0^1 ((\epsilon\gamma_l)^2 + |w - s_l|^2(t))^{p_n/2} dt\right)^{1/p_n - 1} \\ &\cdot \frac{p_n}{2}\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t)\right)^{p_n/2-1} \cdot 2\left\langle (w - s_l)(t), h(t)\right\rangle dt. \end{aligned} \tag{4.29}$$

Applying Hölder's inequality with coefficients $\frac{p_n}{p_n-2k}$ and $\frac{p_n}{2}$ and Cauchy-Schwarz inequality, we obtain that, for any $k = 1, 2, 3$,

$$\begin{aligned}
&\left|\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t)\right)^{p_n/2-k} \langle (w-s_l)(t), h_1(t)\rangle \cdots \langle (w-s_l)(t), h_k(t)\rangle \, dt\right| \\
\leq &\left(\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t)\right)^{p_n/2} dt\right)^{1-2k/p_n} \\
&\cdot \left(\int_0^1 |w - s_l|^{p_n/2}(t) |h_1|^{p_n/(2k)}(t) \cdots |h_k|^{p_n/(2k)}(t) dt\right)^{2k/p_n} \\
\leq &\left(\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t)\right)^{p_n/2} dt\right)^{1-2k/p_n} \cdot \left(\int_0^1 |w - s_l|^{p_n}(t) dt\right)^{k/p_n} \prod_{i=1}^k \|h_i\|_{p_n}.
\end{aligned} \tag{4.30}$$

Applying (4.30) for $k = 1$, together with (4.29), we get

$$|Df(w)[h]| \leq \frac{1}{\eta_n} \left(\frac{\int_0^1 |w - s_l|^{p_n}(t) dt}{\int_0^1 ((\epsilon\gamma_l)^2 + |w - s|^2(t))^{p_n/2} dt}\right)^{1/p_n} \|h\|_{p_n} \leq \frac{\|h\|_\infty}{\eta_n}$$

and so

$$\sup_{w \in D^p} \|Df(w)\| \leq \frac{1}{\eta_n}. \tag{4.31}$$

Step 2: Bounding the second derivative of f of (4.28)

Note that

$$D^2 f(w)[h_1, h_2] = A + B \tag{4.32}$$

for

$$\begin{aligned}
A = &\frac{1}{\eta_n} \left[\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t)\right)^{p_n/2-1} \cdot \langle (w - s_l)(t), h_2(t)\rangle \, dt\right] \\
&\cdot \frac{1 - p_n}{p_n} \left[\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t)\right)^{p_n/2} dt\right]^{1/p_n - 2} \\
&\cdot \frac{p_n}{2} \int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t)\right)^{p_n/2-1} \cdot 2\langle (w - s_l)(t), h_1(t)\rangle \, dt \\
= &\frac{1 - p_n}{\eta_n} \prod_{i=1}^2 \left\{\left[\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t)\right)^{p_n/2-1} \cdot \langle (w - s_l)(t), h_i(t)\rangle \, dt\right]\right\} \\
&\cdot \left[\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t)\right)^{p_n/2} dt\right]^{1/p_n - 2} \\
B = &\frac{1}{\eta_n} \left[\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t)\right)^{p_n/2} dt\right]^{1/p_n - 1}
\end{aligned}$$

$$\cdot\left[\int_0^1 \frac{p_n-2}{2}\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2-2}\cdot 2\left\langle (w-s_l)(t),h_1(t)\right\rangle\left\langle (w-s_l)(t),h_2(t)\right\rangle dt\right.$$

$$\left.+\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2-1}\left\langle h_1(t),h_2(t)\right\rangle dt\right]. \tag{4.33}$$

Notice that, by (4.30) with $k=1$,

$$|A|\leq\frac{p_n-1}{\eta_n}\left(\frac{\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^2}{\left(\int_0^1((\epsilon\gamma_l)^2+|w-s_l|^2(t))^{p_n/2}dt\right)^3}\right)^{1/p_n}\|h_1\|_{p_n}\|h_2\|_{p_n}. \tag{4.34}$$

Furthermore, by Hölder's inequality with coefficients $\frac{p_n}{p_n-2}$ and $\frac{p_n}{2}$ and by the Cauchy-Schwarz inequality,

$$\begin{aligned}&\left|\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2-1}\left\langle h_1(t),h_2(t)\right\rangle dt\right|\\ \leq&\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}\right)^{1-2/p_n}\left(\int_0^1\left\langle h_1(t),h_2(t)\right\rangle^{p_n/2}\right)^{2/p_n}\\ \leq&\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}\right)^{1-2/p_n}\|h_1\|_{p_n}\|h_2\|_{p_n}.\end{aligned} \tag{4.35}$$

By (4.30) and (4.35),

$$\begin{aligned}|B|\leq&\frac{p_n-2}{\eta_n}\left(\frac{\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^2}{\left(\int_0^1((\epsilon\gamma_l)^2+|w-s_l|^2(t))^{p_n/2}dt\right)^3}\right)^{1/p_n}\|h_1\|_{p_n}\|h_2\|_{p_n}\\ &+\frac{1}{\eta_n}\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right)^{-1/p_n}\|h_1\|_{p_n}\|h_2\|_{p_n}.\end{aligned} \tag{4.36}$$

By (4.32), (4.34) and (4.36),

$$\begin{aligned}&|D^2f(w)[h_1,h_2]|\\ \leq&\left[\frac{2p_n-3}{\eta_n}\left(\frac{\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^2}{\left(\int_0^1((\epsilon\gamma_l)^2+|w-s_l|^2(t))^{p_n/2}dt\right)^3}\right)^{1/p_n}\right.\\ &\left.+\frac{1}{\eta_n}\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right)^{-1/p_n}\right]\|h_1\|_{p_n}\|h_2\|_{p_n}\\ =&\frac{1}{\eta_n}\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right)^{-1/p_n}\\ &\cdot\left[(2p_n-3)\left(\frac{\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^2}{\left(\int_0^1((\epsilon\gamma_l)^2+|w-s_l|^2(t))^{p_n/2}dt\right)^2}\right)^{1/p_n}+1\right]\|h_1\|_{p_n}\|h_2\|_{p_n}\end{aligned}$$

$$\leq \frac{2p_n - 2}{\eta_n} \left(\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t) \right)^{p_n/2} dt \right)^{-1/p_n} \|h_1\|_{p_n} \|h_2\|_{p_n}$$

$$\leq \frac{2p_n - 2}{\eta_n(\epsilon\gamma_l)} \|h_1\|_\infty \|h_2\|_\infty$$

and so

$$\sup_{w \in D^p} \|D^2 f(w)\| \leq 2\frac{p_n - 1}{\eta_n(\epsilon\gamma_l)}. \tag{4.37}$$

Step 3: Bounding the third derivative of f of (4.28)

Finally,

$$D^3 f(w)[h_1, h_2, h_3] = C + D, \tag{4.38}$$

where C comes from differentiating A of (4.33) and is given by

$$C = E + F$$

for

$$\begin{aligned} E =& \frac{1 - p_n}{\eta_n} \sum_{1 \leq i \neq j \leq 2} \left\{ \int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t) \right)^{p_n/2-1} \langle (w - s_l)(t), h_i(t) \rangle \, dt \right. \\ & \cdot \int_0^1 \left[\frac{p_n - 2}{2} \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t) \right)^{p_n/2-2} \langle (w - s_l)(t), h_j(t) \rangle \cdot 2 \langle (w - s_l)(t), h_3(t) \rangle \right. \\ & \left. + \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t) \right)^{p_n/2-1} \langle h_j(t), h_3(t) \rangle \right] dt \\ & \left. \cdot \left[\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t) \right)^{p_n/2} dt \right]^{1/p_n - 2} \right\} \\ F =& \frac{(1 - p_n)(1 - 2p_n)}{p_n \eta_n} \left\{ \prod_{i=1}^3 \left[\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t) \right)^{p_n/2-1} \cdot \langle (w - s_l)(t), h_i(t) \rangle \, dt \right] \right\} \\ & \cdot \left[\int_0^1 \left((\epsilon\gamma_l)^2 + |w - s_l|^2(t) \right)^{p_n/2} dt \right]^{1/p_n - 3} \end{aligned} \tag{4.39}$$

and D comes from differentiating B of (4.33) and is given by

$$D = G + H$$

for

$$G = \frac{1-p_n}{\eta_n}\left[\int_0^1 \left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right]^{1/p_n-2}$$
$$\cdot \int_0^1 \left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2-1}\langle (w-s_l)(t),h_3(t)\rangle\, dt$$
$$\cdot \left[\int_0^1 (p_n-2)\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2-2}\cdot\langle (w-s_l)(t),h_1(t)\rangle\langle (w-s_l)(t),h_2(t)\rangle\, dt\right.$$
$$\left. + \int_0^1 \left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2-1}\langle h_1(t),h_2(t)\rangle\, dt\right]$$
$$H = \frac{p_n-2}{\eta_n}\left[\int_0^1 \left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right]^{1/p_n-1}$$
$$\cdot\left\{\int_0^1\left[(p_n-2)\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2-2}\sum_{\substack{1\leq i,j,k\leq 3\\ i,j,k \text{ distinct}}}\langle (w-s_l)(t),h_i(t)\rangle\langle h_j(t),h_k(t)\rangle\right]dt\right.$$
$$\left.+(p_n-4)\int_0^1\left[\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2-3}\prod_{i=1}^3\langle (w-s_l)(t),h_i(t)\rangle\right]dt\right\}. \tag{4.40}$$

So

$$D^3f(w)[h_1,h_2,h_3] = E+F+G+H \tag{4.41}$$

for E,F,G,H defined by (4.39) and (4.40). By (4.30) and (4.35),

$$|E| \leq \frac{2(p_n-1)\|h_1\|_{p_n}\|h_2\|_{p_n}\|h_3\|_{p_n}}{\eta_n}$$
$$\cdot\left(\frac{(p_n-2)\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^{3/p_n}}{\left(\int_0^1((\epsilon\gamma_l)^2+|w-s_l|^2(t))^{p_n/2}dt\right)^{5/p_n}}+\frac{\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^{1/p_n}}{\left(\int_0^1((\epsilon\gamma_l)^2+|w-s_l|^2(t))^{p_n/2}dt\right)^{3/p_n}}\right)$$
$$|F| \leq \frac{(p_n-1)(2p_n-1)\|h_1\|_{p_n}\|h_2\|_{p_n}\|h_3\|_{p_n}}{p_n\eta_n}\cdot\frac{\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^{3/p_n}}{\left(\int_0^1((\epsilon\gamma_l)^2+|w-s_l|^2(t))^{p_n/2}dt\right)^{5/p_n}}$$
$$|G| \leq \frac{(p_n-1)\|h_1\|_{p_n}\|h_2\|_{p_n}\|h_3\|_{p_n}}{\eta_n}$$
$$\cdot\left(\frac{(p_n-2)\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^{3/p_n}}{\left(\int_0^1((\epsilon\gamma_l)^2+|w-s_l|^2(t))^{p_n/2}dt\right)^{5/p_n}}+\frac{\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^{1/p_n}}{\left(\int_0^1((\epsilon\gamma_l)^2+|w-s_l|^2(t))^{p_n/2}dt\right)^{3/p_n}}\right)$$
$$|H| \leq \frac{(p_n-2)\|h_1\|_{p_n}\|h_2\|_{p_n}\|h_3\|_{p_n}}{\eta_n}$$

$$\cdot\left(\frac{(p_n-4)\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^{3/p_n}}{\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right)^{5/p_n}}+\frac{6\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^{1/p_n}}{\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right)^{3/p_n}}\right),\tag{4.42}$$

where the inequality for $|H|$ uses the following bound obtained by applying Hölder's inequality with coefficients $\frac{p_n}{p_n-4}$ and $\frac{p_n}{4}$ and Cauchy-Schwarz inequality

$$\begin{aligned}&\left|\int_0^1\left[\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2-2}\sum_{\substack{1\le i,j,k\le 3\\ i,j,k\text{ distinct}}}\langle(w-s_l)(t),h_i(t)\rangle\langle h_j(t),h_k(t)\rangle\right]dt\right|\\
&\le\sum_{\substack{1\le i,j,k\le 3\\ i,j,k\text{ distinct}}}\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right)^{1-4/p_n}\left(\int_0^1|w-s_l|^{p_n/4}(t)\prod_{i=1}^3|h_i|^{p_n/4}(t)dt\right)^{4/p_n}\\
&\le\sum_{\substack{1\le i,j,k\le 3\\ i,j,k\text{ distinct}}}\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right)^{1-4/p_n}\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^{1/p_n}\prod_{i=1}^3\|h_i\|_{p_n}.\end{aligned}$$

By (4.41) and (4.42),

$$\begin{aligned}&|D^3f(w)[h_1,h_2,h_3]|\\
\le&\frac{6p_n^2\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^{3/p_n}\|h_1\|_{p_n}\|h_2\|_{p_n}\|h_3\|_{p_n}}{\eta_n\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right)^{5/p_n}}\\
&+\frac{9p_n\left(\int_0^1|w-s_l|^{p_n}(t)dt\right)^{1/p_n}\|h_1\|_{p_n}\|h_2\|_{p_n}\|h_3\|_{p_n}}{\eta_n\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right)^{3/p_n}}\\
\le&\frac{15p_n^2\|h_1\|_{p_n}\|h_2\|_{p_n}\|h_3\|_{p_n}}{\eta_n}\left(\int_0^1\left((\epsilon\gamma_l)^2+|w-s_l|^2(t)\right)^{p_n/2}dt\right)^{-2/p_n}\\
\le&\frac{15p_n^2}{(\epsilon\gamma_l)^2\eta_n}\|h_1\|_\infty\|h_2\|_\infty\|h_3\|_\infty\end{aligned}$$

and so

$$\|D^3f(w)\|\le\frac{15p_n^2}{(\epsilon\gamma_l)^2\eta_n}.\tag{4.43}$$

Step 4: Combining the bounds

The result now follows by combining (4.31), (4.37) and (4.43). Indeed, note that, by the chain rule,

$$\begin{aligned}
&D^3 g_{l,n}(w)[h_1,h_2,h_3]\\
=&\phi'''\left(f(w)-\frac{\gamma_l\sqrt{1+\epsilon^2}}{\eta_n}\right)\cdot\prod_{i=1}^{3}Df(w)[h_i]\\
&+\phi''\left(f(w)-\frac{\gamma_l\sqrt{1+\epsilon^2}}{\eta_n}\right)\cdot\sum_{\substack{1\leq i,j,k\leq 3\\ i,j,k\text{distinct}}}D^2f(w)[h_i,h_j]Df(w)[h_k]\\
&+\phi'\left(f(w)-\frac{\gamma_l\sqrt{1+\epsilon^2}}{\eta_n}\right)D^3f(w)[h_1,h_2,h_3].
\end{aligned}$$

By (4.31), (4.37) and (4.43) and the fact that ϕ',ϕ'',ϕ''' are all bounded, we get that, for all $w\in D^p$,

$$\|D^3 g_{l,n}(w)\|\leq C_3p_n^2\eta_n^{-3},$$

for some constant C_3. Similar bounds may be obtained for the first and second derivative of $g_{l,n}$:

$$\|Dg_{l,n}(w)\|\leq C_1\eta_n^{-1},\quad \|D^2g_{l,n}(w)\|\leq C_2p_n\eta_n^{-1},$$

for constants C_1,C_2. Since ϕ is also bounded, the product rule yields the desired bound. □

Now, we prove the following result:

Lemma 4.10. *For the set B fixed at the beggining of this Appendix,*

$$\limsup_{n\to\infty}\mathbb{P}[\mathbf{Y}_n\in B]\leq\mathbb{P}[\mathbf{Z}\in B]\quad \textit{and}\quad \liminf n\to\infty\mathbb{P}[\mathbf{Y}_n\in B]\geq\mathbb{P}[\mathbf{Z}\in B].$$

Proof.

Step 1: Proving the first inequality

Note that

$$\begin{aligned}
&\mathbf{Y}_n\in B_l\Longrightarrow\|\mathbf{Y}_n-s_l\|<\gamma_l\Longrightarrow\sup_{t\in[0,1]}\sum_{i=1}^{p}\left((\mathbf{Y}_n(t)-s_l(t))^{(i)}\right)^2<\gamma_l^2\\
&\Longrightarrow\sup_{t\in[0,1]}\left[\sum_{i=1}^{p}\left((\mathbf{Y}_n(t)-s_l(t))^{(i)}\right)^2+(\epsilon\gamma_l)^2\right]<\gamma_l^2(1+\epsilon^2)\\
&\Longrightarrow\left\|\sqrt{(\epsilon\gamma_l)^2+\sum_{i=1}^{p}\left((\mathbf{Y}_n-s_l)^{(i)}\right)^2}\right\|_{p_n}\leq\gamma_l\sqrt{1+\epsilon^2}\Longrightarrow g_{l,n}(\mathbf{Y}_n)=1.
\end{aligned}$$

Therefore, for all l,

$$\mathbb{1}_{[\mathbf{Y}_n\in B_l]}\leq g_{l,n}(\mathbf{Y}_n).\tag{4.44}$$

Also, note that, by Minkowski's inequality and the triangle inequality for the Euclidean norm:

$$\left\|\sqrt{(\epsilon\gamma_l)^2+\sum_{i=1}^p\left((\mathbf{Z}-s_l)^{(i)}\right)^2}\right\|_{p_n}\leq\left\|\sqrt{(\epsilon\gamma_l)^2+\sum_{i=1}^p\left((\mathbf{Z}_n-s_l)^{(i)}\right)^2}\right\|_{p_n}+\|\mathbf{Z}_n-\mathbf{Z}\|.$$

Therefore, if $\|\mathbf{Z}-s_l\|>\gamma_l$ then as $p_n\xrightarrow{n\to\infty}\infty$:

$$\begin{aligned}&\liminf_{n\to\infty}\left\|\sqrt{(\epsilon\gamma_l)^2+\sum_{i=1}^p\left((\mathbf{Z}_n-s_l)^{(i)}\right)^2}\right\|_{p_n}\\ &\geq\liminf_{n\to\infty}\left\{\left\|\sqrt{(\epsilon\gamma_l)^2+\sum_{i=1}^p\left((\mathbf{Z}-s_l)^{(i)}\right)^2}\right\|_{p_n}-\|\mathbf{Z}_n-\mathbf{Z}\|\right\}\\ &=\sup_{t\in[0,1]}\sqrt{(\epsilon\gamma_l)^2+\sum_{i=1}^p\left((\mathbf{Z}(t)-s_l(t))^{(i)}\right)^2}>\gamma_l(1+\epsilon^2)^{1/2}.\end{aligned}$$

This, means that, if $p_n\xrightarrow{n\to\infty}\infty$, $\|\mathbf{Z}-s_l\|>\gamma_l$ and $\eta_n\xrightarrow{n\to\infty}0$ then $g_{l,n}(\mathbf{Z}_n)=0$ for sufficiently large n, i.e.

$$g_{l,n}(\mathbf{Z}_n)\leq\mathbb{1}_{[\|\mathbf{Z}-s_l\|\leq\gamma_l]},\quad\text{as long as } p_n\xrightarrow{n\to\infty}\infty,\eta_n\xrightarrow{n\to\infty}0\text{ and } n\text{ is large.}\tag{4.45}$$

By those properties, taking $p_n\to\infty$ and $\eta_n\to0$ such that $\tau_n\eta_n^{-3}p_n^2\to0$, we obtain:

$$\begin{aligned}\limsup_{n\to\infty}\mathbb{P}[\mathbf{Y}_n\in B]&\overset{(4.44)}{\leq}\limsup_{n\to\infty}\mathbb{E}\left[\prod_{l=1}^L g_{l,n}(\mathbf{Y}_n)\right]\\ &\overset{(2.4)}{\leq}\limsup_{n\to\infty}\left\{\mathbb{E}\left[\prod_{l=1}^L g_{l,n}(\mathbf{Z}_n)\right]+C\tau_n\left\|\prod_{l=1}^L g_{l,n}\right\|_{M^0}\right\}\\ &\overset{\text{Fatou},(4.27)}{\leq}\mathbb{E}\left[\limsup_{n\to\infty}\prod_{l=1}^L g_{l,n}(\mathbf{Z}_n)\right]\overset{(4.45)}{\leq}\mathbb{P}[\mathbf{Z}\in B].\end{aligned}$$

Step 2: Proving the second inequality

We define:

$$g_{l,n}^*(w)=\phi\left(\frac{\left\|\sqrt{(\epsilon\gamma_l)^2+\sum_{i=1}^p\left((w-s_l)^{(i)}\right)^2}\right\|_{p_n}-\gamma_l\sqrt{\epsilon^2+(1-\theta)^2}(\delta\wedge\frac{r_n}{2})^{1/p_n}+\eta_n}{\eta_n}\right)$$

for $\theta > 0$ fixed and $\delta > 0$ such that:

$$\forall n \in \mathbb{N}: \quad \|\mathbf{Y}_n - s_l\| \geq \gamma_l \Longrightarrow \text{leb}\{t : |\mathbf{Y}_n(t) - s_l(t)| \geq \gamma_l(1-\theta)\} \geq \left(\delta \wedge \frac{r_n}{2}\right),$$

where leb denotes the Lebesgue measure. Such a δ exists for the following reason. The collection $(s_l, 1 \leq l \leq L)$ is uniformly equicontinuous and $\mathbf{Y}_n$ are constant on intervals of length at least r_n. The $\delta > 0$ we choose is such that:

$$|t_1 - t_2| \leq \delta \quad \Longrightarrow \quad |s_l(t_1) - s_l(t_2)| \leq \frac{\theta\gamma_l}{2}.$$

If $\|\mathbf{Y}_n - s_l\| \geq \gamma_l$ then $|\mathbf{Y}_n(t_0) - s_l(t_0)| > \gamma_l\left(1 - \frac{\theta}{2}\right)$ for some t_0. Then, there exists an interval I_0 with t_0 being one of its endpoints and of length $\frac{r_n}{2} \wedge \delta$, such that $\mathbf{Y}_n$ is constant on I_0 and $|s_l(t) - s_l(t_0)| \leq \frac{\theta\gamma_l}{2}$ for all $t \in I_0$. Then, for $t \in I_0$ we obtain:

$$\begin{aligned}&|\mathbf{Y}_n(t) - s_l(t)| \geq |\mathbf{Y}_n(t_0) - s_l(t_0)| - |\mathbf{Y}_n(t_0) - \mathbf{Y}_n(t)| - |s_l(t) - s_l(t_0)| \\ &\geq \left(1 - \frac{\theta}{2}\right)\gamma_l - \frac{\theta\gamma_l}{2} = \gamma_l(1-\theta).\end{aligned}$$

It follows that:

$$\|\mathbf{Y}_n - s_l\| \geq \gamma_l \Longrightarrow \left\|\sqrt{\sum_{i=1}^p \left((\mathbf{Y}_n - s_l)^{(i)}\right)^2}\right\|_{p_n} \geq \gamma_l(1-\theta)\left(\delta \wedge \frac{r_n}{2}\right)^{1/p_n} \Longrightarrow$$

$$\left\|\sqrt{(\epsilon\gamma_l)^2 + \sum_{i=1}^p \left((\mathbf{Y}_n - s_l)^{(i)}\right)^2}\right\|_{p_n} \geq \gamma_l\sqrt{\epsilon^2 + (1-\theta)^2}\left(\delta \wedge \frac{r_n}{2}\right)^{1/p_n} \Longrightarrow g^*_{l,n}(\mathbf{Y}_n) = 0.$$

Therefore, for all l:

$$\mathbb{1}_{[\mathbf{Y}_n \in B_l]} \geq g^*_l(\mathbf{Y}_n). \tag{4.46}$$

Also, again, it can be shown that for any finite L and $\gamma := \min_{1\leq l\leq L}\gamma_l$:

$$\left\|\prod_{l=1}^L g^*_{l,n}\right\|_{M^0} \leq Cp_n^2(\epsilon\gamma)^{-2}\eta_n^{-3} \quad \text{for some constant } C \text{ independent of } p_n, \epsilon, \gamma \text{ and } \eta_n. \tag{4.47}$$

Now suppose $\eta_n \to 0$, $p_n \to \infty$ and $r_n^{1/p_n} \to 1$. Also suppose that $\|\mathbf{Z} - s_l\| < \gamma_l(1-\theta)$ so that there exists $\alpha > 0$ such that a.s. $\|\mathbf{Z}_n - s_l\| < \gamma_l(1-\theta) - \alpha$ for n large enough. Then, for large n:

$$\begin{aligned}\left\|\sqrt{(\epsilon\gamma_l)^2 + \sum_{i=1}^p \left((\mathbf{Z}_n - s_l)^{(i)}\right)^2}\right\|_{p_n} &\leq \sqrt{(\epsilon\gamma_l)^2 + \|\mathbf{Z}_n - s_l\|^2} \leq \gamma_l\sqrt{\epsilon^2 + (1-\theta-\alpha\gamma_l^{-1})^2} \\ &< \gamma_l\sqrt{\epsilon^2 + (1-\theta)^2}\left(\delta \wedge \frac{r_n}{2}\right)^{1/p_n} - \eta_n\end{aligned}$$

because $\left(\delta \wedge \frac{r_n}{2}\right)^{1/p_n} \xrightarrow{n\to\infty} 1$ and $\eta_n \xrightarrow{n\to\infty} 0$. So if $\eta_n \to 0$, $p_n \to \infty$ and $r_n^{1/p_n} \to 1$ then:

$$\|\mathbf{Z} - s_l\| \leq \gamma_l(1-\theta) \Longrightarrow g^*_{l,n}(\mathbf{Z}_n) = 1$$

for n large enough, i.e.:

$$\mathbb{1}_{[\|\mathbf{Z}-s_l\|<\gamma_l(1-\theta)]} \leq g^*_{l,n}(\mathbf{Z}_n). \tag{4.48}$$

Let $\eta_n \to 0$ and $p_n \to \infty$ be such that $r_n^{1/p_n} \to 1$ and $\tau_n p_n^2 \eta_n^{-3} \to 0$. This is possible by the assumption that $\tau_n \log^2(1/r_n) \to 0$. Indeed, having $r_n^{1/p_n} \to 1$, all we require is that $\log\left(r_n^{1/p_n}\right)\eta_n^3 \to 0$ slower than $\tau_n \log^2(1/r_n) \to 0$, because then:

$$\tau_n p_n^2 \eta_n^{-3} = \frac{\tau_n \left(\log(r_n)\right)^2}{\left(\frac{1}{p_n}\log\left(r_n\right)\right)^2 \eta_n^3} = \frac{\tau_n \left(\log(1/r_n)\right)^2}{\left(\log\left(r_n^{1/p_n}\right)\right)^2 \eta_n^3} \to 0$$

For instance, if $r_n \to 0$ and $\tau_n \to 0$, we require p_n and η_n to be such that $\frac{\eta_n^3}{\tau_n} \to \infty$ and $p_n^2 \to \infty$ faster than $(\log r_n)^2$ but slower than $\frac{\eta_n^3}{\tau_n}$.
Then:

$$\begin{aligned}
\liminf_{n\to\infty} \mathbb{P}\left[\mathbf{Y}_n \in B\right] &\overset{(4.46)}{\geq} \liminf_{n\to\infty} \mathbb{E}\left[\prod_{l=1}^{L} g^*_{l,n}(\mathbf{Y}_n)\right] \\
&\overset{(2.4)}{\geq} \liminf_{n\to\infty} \left\{\mathbb{E}\left[\prod_{l=1}^{L} g^*_{l,n}(\mathbf{Z}_n)\right] - C\tau_n \left\|\prod_{l=1}^{L} g^*_{l,n}\right\|_{M^0}\right\} \\
&\overset{\text{Fatou},(4.47)}{\geq} \mathbb{E}\left[\liminf_{n\to\infty} \prod_{l=1}^{L} g^*_{l,n}(\mathbf{Z}_n)\right] \\
&\overset{(4.48)}{\geq} \mathbb{P}\left[\bigcap_{1\leq l\leq L} \left(\|\mathbf{Z}-s_l\| < \gamma_l(1-\theta)\right)\right].
\end{aligned}$$

Since the choice of $\theta \in (0,1)$ was arbitrary, we conclude that:
$\liminf_{n\to\infty} \mathbb{P}\left[\mathbf{Y}_n \in B\right] \geq \mathbb{P}(\mathbf{Z} \in B)$. □

Lemma 4.10 now implies that, for any set B described at the beginning of this Appendix, $\mathbb{P}[\mathbf{Y}_n \in B] \xrightarrow{n\to\infty} \mathbb{P}[\mathbf{Z} \in B]$, which finishes the proof of Proposition 2.3.

arXiv: 1710.09263

Multivariate functional approximations via Stein's method of exchangeable pairs

Mikołaj J. Kasprzak,

Mikołaj Kasprzak
Department of Statistics
University of Oxford
24-29 St Giles'
Oxford, OX1 3LB
United Kingdom
e-mail: kasprzak@stats.ox.ac.uk

Abstract: We combine the multivariate method of exchangeable pairs with Stein's method for functional approximation and give a general linearity condition under which an abstract Gaussian approximation theorem for stochastic processes holds. We apply this approach to estimate the distance from a pre-limiting mixture process of a sum of random variables chosen from an array according to a random permutation and prove a functional combinatorial central limit theorem. We also consider a graph-valued process and bound the speed of convergence of the joint distribution of its rescaled edge and two-star counts to a two-dimensional continuous Gaussian process.

MSC 2010 subject classifications: Primary 60B10, 60F17; secondary 60B12, 60J65, 60E05, 60E15.
Keywords and phrases: Stein's method, functional convergence, exchangeable pairs, multivariate processes.

1. Introduction

In [Ste72] Stein observed that a random variable Z has standard normal law if and only if

$$\mathbb{E}Zf(Z) = \mathbb{E}f'(Z) \tag{1.1}$$

for all smooth functions f. Therefore, if, for a random variable W with mean 0 and variance 1,

$$|\mathbb{E}f'(W) - \mathbb{E}Wf(W)| \tag{1.2}$$

is close to zero for a large class of functions f, then the law of W should be approximately Gaussian. In [Ste86], Stein combined this observation with his *exchangeable-pair* approach. Therein, for a centred and scaled random variable W, its copy W' is constructed in such a way that (W, W') forms an exchangeable pair and the linear regression condition:

$$\mathbb{E}\left[W' - W|W\right] = -\lambda W \tag{1.3}$$

is satisfied for some $\lambda > 0$. This, in many cases, simplifies the process of obtaining bounds on the distance of W from the normal distribution.

This approach was extended in [RR97] to examples in which an approximate linear regression condition holds:

$$\mathbb{E}\left[W' - W|W\right] = -\lambda W + R$$

for some remainder R. A multivariate version of the method was first described in [CM08] and then in [RR09]. In [RR09], for an exchangeable pair of d-dimensional vectors (W, W') the following condition is used:

$$\mathbb{E}[W' - W|W] = -\Lambda W + R$$

for some invertible matrix Λ and a remainder term R. The approach of [RR09] was further reinterpreted and combined with the approach of [CM08] in [Mec09].

On the other hand, in the seminal paper [Bar90], Barbour addressed the problem of providing bounds on the rate of convergence in functional limit results (or invariance principles as they are often called in the literature). He observed that Stein's logic of [Ste72] may also be used in the setup of the Functional Central Limit Theorem. He found a condition, similar to (1.1), characterising the distribution of a standard real Wiener process. Combined with Taylor's theorem, it allowed Barbour to obtain a bound on the rate of convergence in the celebrated Donsker's invariance principle.

This paper is the first attempt to combine the method of exchangeable pairs with functional approximations. We do so in the context of multivariate processes and provide a novel approach to bounding their distances from Gaussian processes. Our approach is influenced by the setup of [RR09] and [Bar90].

1.1. Motivation

We are motivated by a number of (finite-dimensional) examples studied in Stein's method literature using exchangeable pairs, which could be extended to the functional setting. Functional limit results play an important role in applied fields. Researchers often choose to model discrete phenomena with continuous processes arising as scaling limits of discrete ones. The reason is that those scaling limits may be studied using stochastic analysis and are more robust to changes in local details. Questions about the rate of convergence in functional limit results are equivalent to ones about the error those researchers make. Obtaining bounds on a certain distance between the scaled discrete and the limiting continuous processes provides a way of quantifying this error.

We consider two main examples. The first one is a combinatorial functional central limit theorem. The second one considers a two-dimensional process representing edge and two-star counts in a graph-valued process created by unveiling subsequent vertices of a Bernoulli random graph as time progresses.

The former is a functional version of the result proved qualitatively in [HC78] and quantitatively in [CF15] and an extension of the main result of [BJ09]. It considers an array $\{X_{i,j} : i, j = 1, \cdots, n\}$ of i.i.d. random variables, which are then used to create a stochastic process:

$$t \mapsto \frac{1}{s_n} \sum_{i=1}^{\lfloor nt \rfloor} X_{i\pi(i)}, \tag{1.4}$$

where s_n is the variance of $\sum_{i=1}^{n} X_{i\pi(i)}$ and π is a uniform random permutation on $\{1, \cdots, n\}$. The motivation for studying this and similar topics comes from permutation tests in non-parametric statistics. Similar setups, yet with a deterministic array of numbers, and in a finite-dimensional context, have also been considered by other authors (see [WW44] for one of the first works on this topic and [Bol84], [Gol05], [NR12] for quantitative results).

The second example, which considers Bernoulli random graphs, goes back to [JN91]. It was first studied using exchangeable pairs in a finite-dimensional context in [RR10], where a random vector whose components represent statistics corresponding to the number of edges, two-stars and triangles is studied. The authors bound its distance from a normal distribution. We consider a functional analogue of this result, concentrating, for simplicity, only on the number of edges and two-stars. Our approach can, however, be also easily extended to encompass the number of triangles. All of those statistics are often of interest in applications, for example, when approximating the clustering coefficient of a network or in conditional uniform graph tests.

1.2. Contribution of the paper

The main achievements of the paper are the following:

(a) An abstract approximation theorem (Theorem 4.1), providing a bound on the distance between a stochastic process $\mathbf{Y}_n$ valued in $\mathbb{R}^p$ for some positive integer p and a Gaussian mixture process. The theorem assumes that the process $\mathbf{Y}_n$ satisfies the linear regression condition

$$Df(\mathbf{Y}_n)[\mathbf{Y}_n] = 2\mathbb{E}\left\{Df(\mathbf{Y}_n)\left[(\mathbf{Y}_n - \mathbf{Y}'_n)\Lambda_n\right]|\mathbf{Y}_n\right\} + R_f,$$

for all functions f in a certain class of test functions, some matrix Λ_n and some random variable $R_f = R_f(\mathbf{Y}_n)$. As noted in Remark 4.5, this condition is an analogue of the condition considered in [RR09]. Theorem 4.1 is used in the derivation of the remaining results of this paper.

(b) A novel functional combinatorial central limit theorem. In Theorem 5.1, we establish a bound on the distance between process (1.4) and a Gaussian mixture, piecewise constant process. Furthermore, a qualitative result showing convergence of process (1.4) to a continuous Gaussian limiting process is provided in Theorem 5.5. Thus, we extend [BJ09], where similar results were proved under the assumption that all the $X_{i,j}$'s for $i, j = 1, \cdots, n$ are deterministic. Our bound is also an extension of [CF15], where a bound on the rate of weak convergence of the law of $\frac{1}{s_n}\sum_{i=1}^{n} X_{i\pi(i)}$ to the standard normal distribution is obtained.

(c) A novel functional limit theorem for statistics corresponding to edge and two-star counts in a Bernoulli random graph, together with a bound on the rate of convergence. We consider a Bernoulli random graph $G(n,p)$ on n vertices with edge probabilities p. Letting $I_{i,j}$, for $i,j=1,\cdots,n$ be the indicator that edge (i,j) is present in the graph, we study a scaled statistic representing the number of edges:

$$\mathbf{T}_n(t) = \frac{\lfloor nt\rfloor - 2}{2n^2}\sum_{i,j=1}^{\lfloor nt\rfloor} I_{i,j}, \quad t\in[0,1]$$

and another one, representing the number of two-stars (i.e. subgraphs which are trees with one internal node and 2 leaves):

$$\mathbf{V}_n(t) = \frac{1}{2n^2}\sum_{\substack{1\leq i,j,k\leq\lfloor nt\rfloor \\ i,j,k \text{ distinct}}} I_{i,j}I_{j,k}, \quad t\in[0,1].$$

Theorem 6.2 provides a bound on the distance between the law of the process

$$t\mapsto \left(\mathbf{T}_n(t)-\mathbb{E}\mathbf{T}_n(t), \mathbf{V}_n(t)-\mathbb{E}\mathbf{V}_n(t)\right), \quad t\in[0,1] \tag{1.5}$$

and the law of a piecewise constant Gaussian process. Theorem 6.4 bounds a distance between the law of (1.5) and the distribution of a two-dimensional continuous Gaussian process.

1.3. Stein's method of exchangeable pairs

The idea behind the exchangeable-pair approach of [Ste86] was the following. In order to obtain a bound on a distance between the distribution of a centred and scaled random variable W and the standard normal law, one can bound (1.2) for functions f coming from a suitable class. Supposing that we can construct a W' such that (W,W') is an exchangeable pair and (1.3) is satisfied, we can write

$$\begin{aligned} 0 =& \mathbb{E}\left[(f(W)+f(W'))(W-W')\right] \\ =& \mathbb{E}\left[(f(W')-f(W))(W-W')\right] + 2\mathbb{E}\left[f(W)\mathbb{E}[W-W'|W]\right] \\ =& \mathbb{E}\left[(f(W')-f(W))(W-W')\right] + 2\lambda\mathbb{E}[Wf(W)]. \end{aligned}$$

It follows that

$$\mathbb{E}[Wf(W)] = \frac{1}{2\lambda}\mathbb{E}\left[(f(W)-f(W'))(W-W')\right].$$

Therefore, using Taylor's theorem,

$$
\begin{aligned}
&|\mathbb{E}[f'(W)] - \mathbb{E}[Wf(W)]| \\
=&\left|\mathbb{E}[f'(W)] + \frac{1}{2\lambda}\mathbb{E}\left[(f(W') - f(W))(W - W')\right]\right| \\
\leq&\left|\mathbb{E}f'(W) - \frac{1}{2\lambda}\mathbb{E}f'(W)(W - W')^2\right| + \frac{\|f''\|_\infty}{2\lambda}\mathbb{E}|W - W'|^3 \\
\leq&\|f'\|_\infty\mathbb{E}\left|\mathbb{E}\left[\frac{1}{2\lambda}(W - W')^2|W\right] - 1\right| + \frac{\|f''\|_\infty}{2\lambda}\mathbb{E}|W - W'|^3 \\
\leq&\frac{\|f'\|_\infty}{2\lambda}\sqrt{\mathrm{Var}\left[\mathbb{E}\left[(W - W')^2|W\right]\right]} + \frac{\|f''\|_\infty}{2\lambda}\mathbb{E}|W - W'|^3,
\end{aligned}
$$

which provides the desired bound.

Before the publication of [CM08, RR09, Mec09] the method was restricted to one-dimensional approximations. It was, however, also used in the context of non-normal approximations (e.g [CDM05, CFR11, Röl07]). More recently several authors have extended and applied the method. Döbler extended it to Beta distribution in [Döb15] and Chen and Fang used it for the combinatorial CLT [CF15].

1.4. Stein's method in its generality

The aim of the general version of Stein's method is to find a bound of the quantity $|\mathbb{E}_{\nu_n}h - \mathbb{E}_\mu h|$, where μ is the target (known) distribution, ν_n is the approximating law and h is chosen from a suitable class of real-valued test functions $\mathcal{H}$. The procedure can be described in terms of three steps. First, an operator $\mathcal{A}$ acting on a class of real-valued functions is sought, such that

$$(\forall f \in \mathrm{Domain}(\mathcal{A}) \quad \mathbb{E}_\nu\mathcal{A}f = 0) \quad \iff \quad \nu = \mu,$$

where μ is our target distribution. Then, for a given function $h \in \mathcal{H}$, the following Stein equation:

$$\mathcal{A}f = h - \mathbb{E}_\mu h$$

is solved. Finally, using properties of the solution and various mathematical tools (among which the most popular are Taylor's expansions in the continuous case, Malliavin calculus, as described in [NP12], and coupling methods), a bound is sought for the quantity $|\mathbb{E}_{\nu_n}\mathcal{A}f_h|$.

Approximations by laws of stochastic processes have not been covered in the Stein's method literature very widely, with the notable exceptions of [Bar90, BJ09, CD13] and recently [Kas17a, Kas17b, BDM18]. [Kas17a, BDM18] establish a method for bounding the speed of weak convergence of continuous-time Markov chains satisfying certain assumptions to diffusion processes. [Kas17b], on the other hand, treats multi-dimensional processes represented by scaled sums of random variables with different dependence structures using Stein's method and establishes bounds on their distances from continuous Gaussian processes.

1.5. Structure of the paper

The paper is organised as follows. In Section 2 we introduce the spaces of test functions which will be used in the main results, and quote results showing that, under certain assumptions, they determine convergence in distribution under the uniform topology. In Section 3 we set up the Stein equation for approximation by a pre-limiting process and provide properties of the solutions. In Section 4 we provide an exchangeable-pair condition and prove an abstract exchangeable-pair-type approximation theorem. Section 5 is devoted to the functional combinatorial central limit theorem example and Section 6 discusses the graph-valued process example.

2. Spaces M, M^1, M^2, M^0

The following notation, similar to the one of [Kas17b], is used throughout the paper. For a function w defined on the interval $[0,1]$ and taking values in a Euclidean space, we define

$$\|w\| = \sup_{t\in[0,1]} |w(t)|,$$

where $|\cdot|$ denotes the Euclidean norm. We also let $D^p = D([0,1],\mathbb{R}^p)$ be the Skorokhod space of all càdlàg functions on $[0,1]$ taking values in $\mathbb{R}^p$. In the sequel, for $i = 1,\cdots,p$, e_i will denote the ith unit vector of the canonical basis of $\mathbb{R}^p$ and the ith component of $x \in \mathbb{R}^p$ will be represented by $x^{(i)}$, i.e. $x = \left(x^{(1)},\cdots,x^{(p)}\right)$. We will often write $\mathbb{E}^W[\,\cdot\,]$ instead of $\mathbb{E}[\,\cdot\,|W]$.

Let $p \in \mathbb{N}$. Let us define:

$$\|f\|_L := \sup_{w\in D^p} \frac{|f(w)|}{1+\|w\|^3},$$

and let L be the Banach space of continuous functions $f : D^p \to \mathbb{R}$ such that $\|f\|_L < \infty$. Following [Bar90], we now let $M \subset L$ consist of the twice Fréchet differentiable functions f, such that:

$$\|D^2 f(w+h) - D^2 f(w)\| \le k_f \|h\|, \tag{2.1}$$

for some constant k_f, uniformly in $w, h \in D^p$. By $D^k f$ we mean the k-th Fréchet derivative of f and the norm of a k-linear form B on L is defined to be $\|B\| = \sup_{\{h:\|h\|=1\}} |B[h,...,h]|$. Note the following lemma, which can be proved in an analogous way to that used to show (2.6) and (2.7) of [Bar90]. We omit the proof here.

Lemma 2.1. *For every $f \in M$, let:*

$$\|f\|_M := \sup_{w\in D^p} \frac{|f(w)|}{1+\|w\|^3} + \sup_{w\in D^p} \frac{\|Df(w)\|}{1+\|w\|^2} + \sup_{w\in D^p} \frac{\|D^2 f(w)\|}{1+\|w\|} + \sup_{w,h\in D^p} \frac{\|D^2 f(w+h) - D^2 f(w)\|}{\|h\|}.$$

Then, for all $f \in M$, we have $\|f\|_M < \infty$.

For future reference, we let $M^1 \subset M$ be the class of functionals $g \in M$ such that:

$$\|g\|_{M^1} := \sup_{w \in D^p} \frac{|g(w)|}{1+\|w\|^3} + \sup_{w \in D^p} \|Dg(w)\| + \sup_{w \in D^p} \|D^2 g(w)\| + \sup_{w,h \in D^p} \frac{\|D^2 f(w+h) - D^2 f(w)\|}{\|h\|} < \infty \tag{2.2}$$

and $M^2 \subset M$ be the class of functionals $g \in M$ such that:

$$\|g\|_{M^2} := \sup_{w \in D^p} \frac{|g(w)|}{1+\|w\|^3} + \sup_{w \in D^p} \frac{\|Dg(w)\|}{1+\|w\|} + \sup_{w \in D^p} \frac{\|D^2 g(w)\|}{1+\|w\|} + \sup_{w,h \in D^p} \frac{\|D^2 f(w+h) - D^2 f(w)\|}{\|h\|} < \infty. \tag{2.3}$$

We also let M^0 be the class of functionals $g \in M$ such that:

$$\|g\|_{M^0} := \sup_{w \in D^p} |g(w)| + \sup_{w \in D^p} \|Dg(w)\| + \sup_{w \in D^p} \|D^2 g(w)\| + \sup_{w,h \in D^p} \frac{\|D^2 f(w+h) - D^2 f(w)\|}{\|h\|} < \infty.$$

We note that $M^0 \subset M^1 \subset M^2 \subset M$. The next proposition is a p-dimensional version of [BJ09, Proposition 3.1] and shows conditions, under which convergence of the sequence of expectations of a functional g under the approximating measures to the expectation of g under the target measure for all $g \in M^0$ implies weak convergence of the measures of interest. Its proof can be found in the appendix of [Kas17b].

Definition 2.2. $\mathbf{Y} \in D([0,1], \mathbb{R}^p)$ *is piecewise constant if* $[0,1]$ *can be divided into intervals of constancy* $[a_k, a_{k+1})$ *such that the Euclidean norm of* $(\mathbf{Y}(t_1) - \mathbf{Y}(t_2))$ *is equal to* 0 *for all* $t_1, t_2 \in [a_k, a_{k+1})$.

Proposition 2.3. *Suppose that, for each* $n \geq 1$*, the random element* $\mathbf{Y}_n$ *of* D^p *is piecewise constant with intervals of constancy of length at least* r_n*. Let* $(\mathbf{Z}_n)_{n \geq 1}$ *be random elements of* D^p *converging weakly in* D^p*, with respect to the Skorokhod topology, to a random element* $\mathbf{Z} \in C([0,1], \mathbb{R}^p)$*. If:*

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{Z}_n)| \leq C\mathscr{T}_n \|g\|_{M^0} \tag{2.4}$$

for each $g \in M^0$ *and if* $\mathscr{T}_n \log^2(1/r_n) \xrightarrow{n \to \infty} 0$*, then* $\mathbf{Y}_n \Rightarrow \mathbf{Z}$ *(converges weakly) in* D^p*, in both the uniform and the Skorokhod topology.*

3. Setting up Stein's method for the pre-limiting approximation

The steps of the construction presented in this section will be similar to those used to set up Stein's method in [Bar90] and [Kas17b]. After defining the process

$\mathbf{D}_n$ whose distribution will be the target measure in Stein's method, we will construct a process $(\mathbf{W}_n(\cdot,u) : u \geq 0)$ for which the target measure is stationary. We will then calculate its infinitesimal generator $\mathcal{A}_n$ and take it as our Stein operator. Next, we solve the Stein equation $\mathcal{A}_n f = g$ using the analysis of [KDV17] and prove some properties of the solution $f_n = \phi_n(g)$, with the most important one being that its second Fréchet derivative is Lipschitz.

3.1. Target measure

Let

$$\mathbf{D}_n(t) = \sum_{i_1,\cdots,i_m=1}^{n} \left(\tilde{Z}^{(1)}_{i_1,\cdots,i_m} J^{(1)}_{i_1,\cdots,i_m}(t), \cdots, \tilde{Z}^{(p)}_{i_1,\cdots,i_m} J^{(p)}_{i_1,\cdots,i_m}(t)\right), \quad t \in [0,1], \tag{3.1}$$

where $\tilde{Z}^{(k)}_{i_1,\cdots,i_m}$'s for $k = 1,\cdots,p$ are centred Gaussian and:

A) the covariance matrix $\Sigma_n \in \mathbb{R}^{(n^m p)\times(n^m p)}$ of the vector $\tilde{Z}$ is positive definite, where $\tilde{Z} \in \mathbb{R}^{(n^m p)}$ is formed out of the $\tilde{Z}^{(k)}_{i_1,\cdots,i_m}$'s in such a way that they appear in the lexicographic order with $\tilde{Z}^{(k)}_{i_1,\cdots,i_m}$ appearing before $\tilde{Z}^{(k+1)}_{j_1,\cdots,j_m}$'s for any $k = 1,\cdots,p-1$ and $i_1,\cdots,i_m,j_1,\cdots,j_m = 1,\cdots,n$;

B) $J^{(k)}_{i_1,\cdots,i_m} \in D([0,1],\mathbb{R})$, for $i_1,\cdots,i_m \in \{1,\cdots,n\}$, $k \in \{1,\cdots,p\}$, are independent of the $\tilde{Z}^{(k)}_{i_1,\cdots,i_m}$'s. A typical example would be $J^{(k)}_{i_1,\cdots,i_m} = \mathbb{1}_{A^{(k)}_{i_1,\cdots,i_m}}$ for some measurable set $A^{(k)}_{i_1,\cdots,i_m}$.

Remark 3.1. *It is worth noting that processes $\mathbf{D}_n$ taking the form (3.1) often approximate interesting continuous Gaussian processes very well. An example is a Gaussian scaled random walk, i.e. $\mathbf{D}_n$ of (3.1), where all the $\tilde{Z}^{(k)}_{i_1,\cdots,i_m}$'s are standard normal and independent, $m = 1$ and $J^{(k)}_i = \mathbb{1}_{[i/n,1]}$ for all $k = 1,\cdots,p$ and $i = 1,\cdots,n$. It approximates Brownian Motion. By Proposition 2.3, under several assumptions, proving by Stein's method that a piece-wise constant process $\mathbf{Y}_n$ is close enough to process $\mathbf{D}_n$ proves $\mathbf{Y}_n$'s convergence in law to the continuous process that $\mathbf{D}_n$ approximates.*

Now let $\{(\mathscr{X}^{(k)}_{i_1,\cdots,i_m}(u), u \geq 0) : i_1,\cdots,i_m = 1,...,n, k = 1,...,p\}$ be an array of i.i.d. Ornstein-Uhlenbeck processes with stationary law $\mathcal{N}(0,1)$, independent of the $J^{(k)}_{i_1,\cdots,i_m}$'s. Consider $\tilde{\mathscr{U}}(u) = (\Sigma_n)^{1/2}\,\mathscr{X}(u)$, where $\mathscr{X}(u) \in \mathbb{R}^{n^m p}$ is formed out of the $\mathscr{X}^{(k)}_{i_1,\cdots,i_m}(u)$'s in such a way that they appear in the same order as the $\tilde{Z}^{(k)}_{i_1,\cdots,i_m}$'s appear in $\tilde{Z}$. Write $\mathscr{U}^{(k)}_{i_1,\cdots,i_m}(u) = \left(\tilde{\mathscr{U}}(u)\right)_{I(k,i_1,\cdots,i_m)}$ using the bijection $I : \{(k,i_1,\cdots,i_m) : i_1,\cdots,i_m = 1,\cdots,n, k = 1,\cdots,p\} \to \{1,\cdots,pn^m\}$, given by:

$$I(k,i_1,\cdots,i_m) = (k-1)n^m + (i_1-1)n^{m-1} + \cdots + (i_{m-1}-1)n + i_m. \tag{3.2}$$

Consider a process:

$$\mathbf{W}_n(t,u) = \left(\mathbf{W}_n^{(1)}(t,u), \cdots, \mathbf{W}_n^{(p)}(t,u)\right), \quad t \in [0,1], u \geq 0,$$

where, for all $k = 1, \cdots, p$:

$$\mathbf{W}_n^{(k)}(t,u) = \sum_{i_1,\cdots,i_m=1}^{n} \mathscr{U}_{i_1,\cdots,i_m}^{(k)}(u) J_{i_1,\cdots,i_m}^{(k)}(t), \quad t \in [0,1], u \geq 0.$$

It is easy to see that the stationary law of the process $(\mathbf{W}_n(\cdot,u))_{u\geq 0}$ is exactly the law of $\mathbf{D}_n$.

3.2. Stein equation

By [Kas17b, Propositions 4.1 and 4.4], the following result is immediate:

Proposition 3.2. *The infinitesimal generator of the process $(\mathbf{W}_n(\cdot,u))_{u\geq 0}$ acts on any $f \in M$ in the following way:*

$$\mathcal{A}_n f(w) = -Df(w)[w] + \mathbb{E}D^2 f(w)\left[\mathbf{D}_n^{(2)}\right].$$

Moreover, for any $g \in M$ such that $\mathbb{E}g(\mathbf{D}_n) = 0$, the Stein equation $\mathcal{A}_n f_n = g$ is solved by:

$$f_n = \phi_n(g) = -\int_0^\infty T_{n,u} g du, \tag{3.3}$$

where $(T_{n,u}f)(w) = \mathbb{E}\left[f(we^{-u} + \sqrt{1-e^{-2u}}\mathbf{D}_n(\cdot)\right]$ Furthermore, for $g \in M$:

$$A)\quad \|D\phi_n(g)(w)\| \leq \|g\|_M\left(1 + \frac{2}{3}\|w\|^2 + \frac{4}{3}\mathbb{E}\|\mathbf{D}_n\|^2\right),$$

$$B)\quad \|D^2\phi_n(g)(w)\| \leq \|g\|_M\left(\frac{1}{2} + \frac{\|w\|}{3} + \frac{\mathbb{E}\|\mathbf{D}_n\|}{3}\right),$$

$$C)\quad \frac{\left\|D^2\phi_n(g)(w+h) - D^2\phi_n(g)(w)\right\|}{\|h\|} \leq \sup_{w,h\in D^p} \frac{\|D^2(g+c)(w+h) - D^2(g+c)(w)\|}{3\|h\|},$$

for any constant function $c : D^p \to \mathbb{R}$ and for all $w, h \in D^p$. Moreover, for all $g \in M^1$, as defined in (2.2),

$$A)\quad \|D\phi_n(g)(w)\| \leq \|g\|_{M^1},$$

$$B)\quad \|D^2\phi_n(g)(w)\| \leq \frac{1}{2}\|g\|_{M^1}$$

and for all $g \in M^2$, as defined in (2.3),

$$\|D\phi_n(g)(w)\| \leq \|g\|_{M^2}.$$

4. An abstract approximation theorem

We now present a theorem which provides an expression for a bound on the distance between some process $\mathbf{Y}_n$ and $\mathbf{D}_n$, defined by (3.1), provided that we can find some $\mathbf{Y}_n'$ such that $(\mathbf{Y}_n, \mathbf{Y}_n')$ is an exchangeable pair satisfying an appropriate condition. Our condition (4.1) is similar to that of [RR09, (1.7)], as we explain in Remark 4.5.

Theorem 4.1. *Assume that $(\mathbf{Y}_n, \mathbf{Y}_n')$ is an exchangeable pair of $D([0,1],\mathbb{R}^p)$-valued random vectors such that:*

$$Df(\mathbf{Y}_n)[\mathbf{Y}_n] = 2\mathbb{E}^{\mathbf{Y}_n} Df(\mathbf{Y}_n)\left[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n\right] + R_f, \tag{4.1}$$

where $\mathbb{E}^{\mathbf{Y}_n}[\cdot] := \mathbb{E}[\cdot|\mathbf{Y}_n]$, for all $f \in M$, some $\Lambda_n \in \mathbb{R}^{p\times p}$ and some random variable $R_f = R_f(\mathbf{Y}_n)$. Let $\mathbf{D}_n$ be defined by (3.1). Then, for any $g \in M$:

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{D}_n)| \le \epsilon_1 + \epsilon_2 + \epsilon_3$$

where $f = \phi_n(g)$, as defined by (3.3), and

$$\begin{aligned}
\epsilon_1 &= \frac{\|g\|_M}{6}\mathbb{E}\|(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n\|\|\mathbf{Y}_n - \mathbf{Y}_n'\|^2,\\
\epsilon_2 &= \left|\mathbb{E}D^2 f(\mathbf{Y}_n)\left[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n, \mathbf{Y}_n - \mathbf{Y}_n'\right] - \mathbb{E}D^2 f(\mathbf{Y}_n)\left[\mathbf{D}_n, \mathbf{D}_n\right]\right|,\\
\epsilon_3 &= |\mathbb{E}R_f|.
\end{aligned}$$

Remark 4.2 (Relevance of terms in the bound)**.** *Term ϵ_1 measures how close $\mathbf{Y}_n$ and $\mathbf{Y}_n'$ are and how small (in a certain sense) Λ_n is. Term ϵ_2 corresponds to the comparison of the covariance structure of $\mathbf{Y}_n - \mathbf{Y}_n'$ and $\mathbf{D}_n$. Estimating this term usually requires some effort yet is possible in several applications (see Theorem 5.1 and 6.2 below). Term ϵ_3 measures the error in the exchangeable-pair linear regression condition (4.1).*

Remark 4.3. *Condition (4.1) is always satisfied, for example with $\Lambda_n = 0$ and $R_f = Df(\mathbf{Y}_n)[\mathbf{Y}_n]$ for all $f \in M$. However, for the bound in Theorem 4.1 to be small, we require the expectation of R_f to be small in absolute value.*

Remark 4.4. *The term*

$$\left|\mathbb{E}D^2 f(\mathbf{Y}_n)\left[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n, \mathbf{Y}_n - \mathbf{Y}_n'\right] - \mathbb{E}D^2 f(\mathbf{Y}_n)\left[\mathbf{D}_n, \mathbf{D}_n\right]\right|$$

in the bound obtained in Theorem 4.1 is an analogue of the second condition in [Mec09, Theorem 3]. Therein, a bound on approximation by $\mathcal{N}(0,\Sigma)$ of a d-dimensional vector X is obtained by constructing an exchangeable pair (X, X') satisfying:

$$\mathbb{E}^X[X' - X] = \Lambda X + E \quad \textit{and} \quad \mathbb{E}^X[(X'-X)(X'-X)^T] = 2\Lambda\Sigma + E'$$

for some invertible matrix Λ and some remainder terms E and E'. In the same spirit, Theorem 4.1 could be rewritten to assume (4.1) and:

$$\mathbb{E}^{\mathbf{Y}_n} D^2 f(\mathbf{Y}_n)\left[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n, \mathbf{Y}_n - \mathbf{Y}_n'\right] = D^2 f(\mathbf{Y}_n)\left[\mathbf{D}_n, \mathbf{D}_n\right] + R_f^1.$$

The bound would then take the form:

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{D}_n)| \leq \frac{\|g\|_M}{6}\mathbb{E}\|(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n\|\|\mathbf{Y}_n - \mathbf{Y}_n'\|^2 + |\mathbb{E}R_f| + |\mathbb{E}R_f^1|.$$

Remark 4.5. *The role of Λ_n in condition (4.1) is equivalent to that played by Λ^{-1} in [RR09] for Λ defined by (1.7) therein. As also observed in [RR09], the condition involving a matrix Λ is a generalisation of the condition of [CM08, Theorem 1], where a scalar is used instead. It should be noted that condition (4.1) is more appropriate in the functional setting than a straightforward adaptation of the condition of [RR09]. This is due to the fact that for general processes $\mathbf{Y}_n$ the properties of the Fréchet derivative do not allow us to treat evaluating the derivative in the direction of $\mathbf{Y}_n - \mathbf{Y}_n'$ as matrix multiplication and multiplying both sides of the hypothetical condition:*

$$-Df(\mathbf{Y}_n)[\Lambda\mathbf{Y}_n] = \mathbb{E}^{\mathbf{Y}_n}Df(\mathbf{Y}_n)[\mathbf{Y}_n - \mathbf{Y}_n']$$

by Λ^{-1} does not give:

$$-Df(\mathbf{Y}_n)[\mathbf{Y}_n] = \mathbb{E}^{\mathbf{Y}_n}Df(\mathbf{Y}_n)[\Lambda^{-1}(\mathbf{Y}_n - \mathbf{Y}_n')].$$

Proof of Theorem 4.1. Our aim is to bound $|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{D}_n)|$ by bounding $|\mathbb{E}\mathcal{A}_n f(\mathbf{Y}_n)|$, where f is the solution to the Stein equation:

$$\mathcal{A}_n f = g - \mathbb{E}g(\mathbf{D}_n),$$

for $\mathcal{A}_n$ defined in Proposition 3.2. Note that, by exchangeability of $(\mathbf{Y}_n, \mathbf{Y}_n')$ and (4.1):

$$\begin{aligned}
0 =&\mathbb{E}\left(Df(\mathbf{Y}_n') + Df(\mathbf{Y}_n)\right)[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n]\\
=&\mathbb{E}\left(Df(\mathbf{Y}_n') - Df(\mathbf{Y}_n)\right)[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n] + 2\mathbb{E}\left\{\mathbb{E}^{\mathbf{Y}_n}Df(\mathbf{Y}_n)\left[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n\right]\right\}\\
=&\mathbb{E}\left(Df(\mathbf{Y}_n') - Df(\mathbf{Y}_n)\right)[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n] + \mathbb{E}Df(\mathbf{Y}_n)[\mathbf{Y}_n] - \mathbb{E}R_f
\end{aligned}$$

and so:

$$\mathbb{E}Df(\mathbf{Y}_n)[\mathbf{Y}_n] = \mathbb{E}\left(Df(\mathbf{Y}_n) - Df(\mathbf{Y}_n')\right)[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n] + \mathbb{E}R_f.$$

Therefore:

$$\begin{aligned}
&|\mathbb{E}\mathcal{A}_n f(\mathbf{Y}_n)|\\
=&\left|\mathbb{E}Df(\mathbf{Y}_n)[\mathbf{Y}_n] - \mathbb{E}D^2f(\mathbf{Y}_n)\left[\mathbf{D}_n, \mathbf{D}_n\right]\right|\\
=&\left|\mathbb{E}\left(Df(\mathbf{Y}_n) - Df(\mathbf{Y}_n')\right)[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n] - \mathbb{E}D^2f(\mathbf{Y}_n)\left[\mathbf{D}_n, \mathbf{D}_n\right] + \mathbb{E}R_f\right|\\
\leq&\left|\mathbb{E}\left(Df(\mathbf{Y}_n) - Df(\mathbf{Y}_n')\right)[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n] - \mathbb{E}D^2f(\mathbf{Y}_n')\left[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n, \mathbf{Y}_n - \mathbf{Y}_n'\right]\right|\\
&+\left|\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n, \mathbf{Y}_n - \mathbf{Y}_n'\right] - \mathbb{E}D^2f(\mathbf{Y}_n)\left[\mathbf{D}_n, \mathbf{D}_n\right]\right| + |\mathbb{E}R_f|\\
\leq&\frac{\|g\|_M}{6}\mathbb{E}\|(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n\|\|\mathbf{Y}_n - \mathbf{Y}_n'\|^2 + |\mathbb{E}R_f|\\
&+\left|\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n, \mathbf{Y}_n - \mathbf{Y}_n'\right] - \mathbb{E}D^2f(\mathbf{Y}_n)\left[\mathbf{D}_n, \mathbf{D}_n\right]\right|,
\end{aligned}$$

where the last inequality follows by Taylor's theorem and Proposition 3.2. □

5. A functional Combinatorial Central Limit Theorem

In this section we consider a functional version of the result proved in [HC78]. Our object of interest is a stochastic process represented by a scaled sum of independent random variables chosen from an $n \times n$ array. Only one random variable is picked from each row and for row i, the corresponding random variable is picked from column $\pi(i)$, where π is a random permutation on $[n] = \{1, \cdots, n\}$. Theorem 5.1 established a bound on the distance between this process and a pre-limiting process and Theorem 5.5 shows convergence of this process, under certain assumptions, to a continuous Gaussian process.

Our analysis in this section is similar to that of [BJ09], where the summands in the scaled sums are chosen from a deterministic array. The authors therein also establish bounds on the approximation by a pre-limit Gaussian process and show convergence to a continuous Gaussian process. Furthermore, they establish a bound on the distance from the continuous Gaussian process for a restricted class of test functions. For random arrays the situation is more involved.

Our setup is analogous to the one considered in [CF15], where a bound on the speed of convergence in the one-dimensional combinatorial central limit theorem is obtained using Stein's method of exchangeable pairs.

5.1. Introduction

Let $\mathbb{X} = \{X_{i,j} : i, j \in [n]\}$ be an $n \times n$ array of independent $\mathbb{R}$-valued random variables, where $n \geq 2$, $\mathbb{E}X_{i,j} = c_{ij}$, $\text{Var}X_{i,j} = \sigma_{ij}^2 \geq 0$ and $\mathbb{E}|X_{i,j}|^3 < \infty$. Suppose that $c_{i\cdot} = c_{\cdot j} = 0$ where $c_{i\cdot} = \sum_{j=1}^n \frac{c_{ij}}{n} = \mathbb{E}X_{i\pi(i)}$, $c_{\cdot j} = \sum_{i=1}^n \frac{c_{ij}}{n}$. Let π be a uniform random permutation of $[n]$, independent of $\mathbb{X}$ and for

$$s_n^2 = \frac{1}{n}\sum_{i,j=1}^n \sigma_{ij}^2 + \frac{1}{n-1}\sum_{i,j=1}^n c_{ij}^2. \tag{5.1}$$

let

$$\mathbf{Y}_n(t) = \frac{1}{s_n}\sum_{i=1}^{\lfloor nt\cdot \rfloor} X_{i\pi(i)} = \frac{1}{s_n}\sum_{i=1}^{n} X_{i\pi(i)}\mathbb{1}_{[i/n,1]}(t), \quad t \in [0,1].$$

We note that $s_n^2 = \text{Var}\left[\sum_{i=1}^n X_{i\pi(i)}\right]$ by the first part of [CF15, Theorem 1.1]. The process $\mathbf{Y}_n$ is similar to the process Y considered in [BJ09] and defined by (1.4) therein with the most important difference being that we allow the $X_{i,j}$'s to be random, whereas the authors in [BJ09] assumed them to be deterministic. Bounds on the distance between one-dimensional distributions of $\mathbf{Y}_n$ and a normal distribution have been obtained via Stein's method in [CF15, Theorem 1.1].

5.2. Exchangeable pair setup

Select uniformly at random two different indices $I, J \in [n]$ and let:

$$\mathbf{Y}_n' = \mathbf{Y}_n - \frac{1}{s_n} X_{I\pi(I)} \mathbb{1}_{[I/n,1]} - \frac{1}{s_n} X_{J\pi(J)} \mathbb{1}_{[J/n,1]} + \frac{1}{s_n} X_{I\pi(J)} \mathbb{1}_{[I/n,1]} + \frac{1}{s_n} X_{J\pi(I)} \mathbb{1}_{[J/n,1]}.$$

Note that $(\mathbf{Y}_n, \mathbf{Y}_n')$ is an exchangeable pair and that for all $f \in M$:

$$\begin{aligned}
&\mathbb{E}^{\mathbf{Y}_n} \left\{ Df(\mathbf{Y}_n) \left[\mathbf{Y}_n - \mathbf{Y}_n' \right] \right\} \\
=&\frac{1}{s_n} \mathbb{E}^{\mathbf{Y}_n} \left\{ Df(\mathbf{Y}_n) \left[X_{I\pi(I)} \mathbb{1}_{[I/n,1]} + X_{J\pi(J)} \mathbb{1}_{[J/n,1]} - X_{I\pi(J)} \mathbb{1}_{[I/n,1]} - X_{J\pi(I)} \mathbb{1}_{[J/n,1]} \right] \right\} \\
=&\frac{1}{n(n-1)s_n} \sum_{i,j=1}^{n} \mathbb{E}^{\mathbf{Y}_n} \left\{ Df(\mathbf{Y}_n) \left[X_{i\pi(i)} \mathbb{1}_{[i/n,1]} + X_{j\pi(j)} \mathbb{1}_{[j/n,1]} \right.\right. \\
&\left.\left. - X_{i\pi(j)} \mathbb{1}_{[i/n,1]} - X_{j\pi(i)} \mathbb{1}_{[j/n,1]} \right] \right\} \\
=&\frac{2}{n-1} Df(\mathbf{Y}_n)[\mathbf{Y}_n] - \frac{2}{n(n-1)s_n} \sum_{i,j=1}^{n} \mathbb{E}^{\mathbf{Y}_n} Df(\mathbf{Y}_n) \left[X_{i,\pi(j)} \mathbb{1}_{[i/n,1]} \right].
\end{aligned}$$

Therefore:

$$\mathbb{E}^{\mathbf{Y}_n} \left\{ Df(\mathbf{Y}_n)[\mathbf{Y}_n - \mathbf{Y}_n'] \right\} = \frac{2}{n-1} Df(\mathbf{Y}_n) \left[\mathbf{Y}_n - \frac{1}{ns_n} \sum_{i,j=1}^{n} \mathbb{E}^{\mathbf{Y}_n} [X_{i,\pi(j)}] \mathbb{1}_{[i/n]} \right]. \tag{5.2}$$

So condition (4.1) is satisfied with

$$\Lambda_n = \frac{n-1}{4} \quad \text{and} \quad R_f = \frac{1}{ns_n} \sum_{i,j=1}^{n} Df(\mathbf{Y}_n) \left[\mathbb{E}^{\mathbf{Y}_n} [X_{i,\pi(j)}] \mathbb{1}_{[i/n]} \right].$$

5.3. Pre-limiting process

Now let $\hat{Z}_i = \frac{1}{\sqrt{n-1}} \sum_{l=1}^{n} X_{il}'' \left(Z_{il} - \frac{1}{n} \sum_{j=1}^{n} Z_{jl} \right)$, for $\mathbb{X}'' = \{X_{ij}'' : i,j \in [n]\}$ being an independent copy of of $\mathbb{X}$ and Z_{il}'s i.i.d. standard normal, independent of all the X_{il}'s and X_{il}'''s. Then, let

$$\mathbf{D}_n(t) = \frac{1}{s_n} \sum_{i=1}^{\lfloor nt \rfloor} \hat{Z}_i, \quad t \in [0,1]. \tag{5.3}$$

We will compare the distribution of $\mathbf{Y}_n$ with the distribution of $\mathbf{D}_n$. $\mathbf{D}_n$ is a conceptually easy process with the same covariance structure as $\mathbf{Y}_n$. It is constructed in a way similar to the process in [BJ09, (3.13)]. Note that $\hat{Z}_i$ has mean 0 for all i and

$$\begin{aligned}
\mathbb{E}\hat{Z}_i^2 =& \frac{1}{n-1}\sum_{l=1}^n \mathbb{E}\left[X_{il}^2\right]\mathbb{E}\left[\left(Z_{il}-\frac{1}{n}\sum_{j=1}^n Z_{jl}\right)^2\right] \\
&+ \frac{1}{n-1}\sum_{1\leq l\neq k\leq n}\mathbb{E}\left[X_{il}X_{ik}\right]\mathbb{E}\left[\left(Z_{il}-\frac{1}{n}\sum_{j=1}^n Z_{jl}\right)\left(Z_{ik}-\frac{1}{n}\sum_{j=1}^n Z_{jk}\right)\right] \\
=& \frac{1}{n-1}\sum_{l=1}^n \mathbb{E}\left[X_{il}^2\right]\left(1-\frac{2}{n}+\frac{1}{n}\right) \\
=& \frac{1}{n}\sum_{l=1}^n \mathbb{E}X_{il}^2 \\
=& \frac{1}{2n^2}\left(2(n-1)\sum_{l=1}^n \mathbb{E}X_{il}^2 + 2\sum_{i=1}^n \mathbb{E}X_{ir}^2\right) \\
=& \frac{1}{2n^2}\left(\sum_{1\leq k\neq l\leq n}\mathbb{E}\left[(X_{ik}-X_{il})^2\right] + 2\sum_{1\leq k\neq l\leq n}\mathbb{E}X_{ik}\mathbb{E}X_{il} + 2\sum_{r=1}^n \mathbb{E}X_{ir}^2\right) \\
=& \frac{1}{2n^2}\left(\sum_{1\leq k\neq l\leq n}\mathbb{E}\left[(X_{ik}-X_{il})^2\right] + 2\sum_{r=1}^n \sigma_{ir}^2\right) \qquad (5.4)
\end{aligned}$$

as $c_{i\cdot}=0$, and, for $i\neq j$,

$$\begin{aligned}
\mathbb{E}\hat{Z}_i\hat{Z}_j =& \frac{1}{n-1}\sum_{k,l=1}^n \mathbb{E}(X_{ik}X_{jl})\mathbb{E}\left[\left(Z_{ik}-\frac{1}{n}\sum_{r=1}^n Z_{rk}\right)\left(Z_{jl}-\frac{1}{n}\sum_{r=1}^n Z_{rl}\right)\right] \\
=& -\frac{1}{n(n-1)}\sum_{k=1}^n c_{ik}c_{jk} \\
=& \frac{1}{2n^2(n-1)}\left(2\sum_{k=1}^n\left(-\mathbb{E}X_{ik}\right)\mathbb{E}X_{jk} - 2(n-1)\sum_{k=1}^n \mathbb{E}X_{ik}\mathbb{E}X_{jk}\right) \\
=& \frac{1}{2n^2(n-1)}\left(2\sum_{1\leq k\neq l\leq n}\mathbb{E}X_{il}\mathbb{E}X_{jk} - 2\sum_{1\leq k\neq l\leq n}\mathbb{E}X_{ik}\mathbb{E}X_{jk}\right) \\
=& \frac{1}{2n^2(n-1)}\sum_{1\leq k\neq l\leq n}\mathbb{E}(X_{ik}-X_{il})(X_{jl}-X_{jk}). \qquad (5.5)
\end{aligned}$$

5.4. Pre-limiting approximation

We have the following theorem, comparing the distribution of $\mathbf{Y}_n$ and $\mathbf{D}_n$:

Theorem 5.1. *For $\mathbf{Y}_n$ defined in Subsection 5.1, $\mathbf{D}_n$ defined in Subsection 5.3 and any $g \in M^1$, as defined in (2.2),*

$$
\begin{aligned}
&|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{D}_n)| \\
\leq & \frac{\|g\|_{M^1}}{n^3(n-1)s_n^3} \sum_{1\leq i,j,k,l,u\leq n} \Big\{ 3\mathbb{E}|X_{ik}|^3 + 5\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|^2 + 7\mathbb{E}|X_{ik}|^2\mathbb{E}|X_{jl}| \\
&+ 5\mathbb{E}|X_{ik}|^2\mathbb{E}|X_{jk}| + 16\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|\mathbb{E}|X_{jl}| + 2\mathbb{E}|X_{iu}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}| \\
&+ 4\mathbb{E}|X_{iu}|\mathbb{E}|X_{il}|\mathbb{E}|X_{jk}| + 6\mathbb{E}|X_{uk}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{jl}| + 2\mathbb{E}|X_{uk}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{jk}| \\
&+ \frac{1}{n}\left(2\mathbb{E}\,|X_{ik}| + 2\mathbb{E}\,|X_{j,l}| + 2\mathbb{E}|X_{uk}| + 2\mathbb{E}|X_{ul}|\right) \sum_{r=1}^n \left(\mathbb{E}|X_{ir}|^2 + |c_{ir}c_{jr}|\right) \Big\} \\
&+ \frac{2\|g\|_{M^1}}{\sqrt{n}} + \frac{4\|g\|_{M^1}}{3ns_n^2} \sum_{i,j=1}^n \sigma_{i,j}^2.
\end{aligned}
$$

Remark 5.2 (Relevance of terms in the bound)**.** *The first long sum in the bound corresponds to ϵ_1 and (to a large extent) ϵ_2 of Theorem 4.1. It represents the usual Berry-Esseen third moment estimate arising as a result of applying Taylor's theorem. Term $\frac{2\|g\|_{M^1}}{\sqrt{n}}$ also comes from the estimation of ϵ_2. The last term corresponds to ϵ_3.*

Remark 5.3. *Assuming that $s_n = O(\sqrt{n})$, we obtain that the bound in Theorem 5.1 is of order $\frac{1}{\sqrt{n}}$.*

Remark 5.4. *If we assume that $\mathbb{E}|X_{ik}|^3 \leq \beta_3$ for all $i, k = 1, \cdots, n$ then the bound simplifies in the following way*

$$
\begin{aligned}
&|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{D}_n)| \\
\leq & \|g\|_{M^1} \left(\frac{58\beta_3 n^2}{(n-1)s_n^3} + \frac{8\beta_3^{1/3}}{n(n-1)s_n^3} \sum_{i,j,r=1}^n |c_{ir}c_{jr}| + \frac{2}{\sqrt{n}} + \frac{4}{3ns_n^2} \sum_{i,j=1}^n \sigma_{i,j}^2 \right).
\end{aligned}
$$

We will use Theorem 4.1 to prove Theorem 5.1. In the proof, in **Step 1**, we justify why Theorem 4.1 may indeed be used in this case. In other words, we check that $\mathbf{D}_n$ of (5.3) satisfies the conditions $\mathbf{D}_n$ of Theorem (4.1) is supposed to satisfy and that the exchangeable-pair condition for $\mathbf{Y}_n$ holds. In **Step 2** we bound terms ϵ_1 and ϵ_3 coming from Theorem 4.1. This is relatively straightforward due to the $\mathbf{Y}_n$ and $\mathbf{Y}_n'$ of Subsection 5.2 being constructed in such a way that they are close to each other and R_f of the same subsection being small. Then, in **Step 3**, we treat the remaining term using a strategy analogous to that of the proof of [BJ09, Theorem 2.1]. The strategy is based on Taylor's expansions and considering copies of $\mathbf{Y}_n$ which are independent of some of the summands in $\mathbf{Y}_n$. Finally, we combine the estimates obtained in the previous steps to obtain the assertion.

Proof of theorem 5.1. We adopt the notation of Subsections 5.1, 5.2 and 5.3.

Furthermore, we fix a function $g \in M^1$, as defined in (2.2) and let $f = \phi_n(g)$, a solution to the Stein equation for $\mathbf{D}_n$, as defined in (3.3).

Step 1. We note that $\mathbf{D}_n$ can be expressed in the following way:

$$\mathbf{D}_n = \sum_{i,l=1}^{n} \left(Z_{il} - \frac{1}{n} \sum_{j=1}^{n} Z_{jl} \right) J_{i,l}, \quad \text{where} \quad J_{i,l}(t) = \frac{X''_{il}}{s_n \sqrt{n-1}} \mathbb{1}_{[i/n,1]}(t),$$

which, together with (5.2), lets us apply Theorem 4.1.

Step 2. For the first term in Theorem 4.1, for any $g \in M^1$:

$$\epsilon_1 = \frac{\|g\|_{M^1}}{6} \mathbb{E}\|(\mathbf{Y}_n - \mathbf{Y}'_n)\Lambda_n\| \|\mathbf{Y}_n - \mathbf{Y}'_n\|^2 \leq \frac{(n-1)\|g\|_{M^1}}{24} \mathbb{E}\|\mathbf{Y}_n - \mathbf{Y}'_n\|^3.$$

We note that:

$$\begin{aligned}
\mathbb{E}\|\mathbf{Y}_n - \mathbf{Y}'_n\|^3 &\leq \frac{8}{s_n^3} \left(\mathbb{E}|X_{I\pi(I)}|^3 + \mathbb{E}|X_{J\pi(J)}|^3 + \mathbb{E}|X_{I\pi(J)}|^3 + \mathbb{E}|X_{J\pi(I)}|^3 \right) \\
&= \frac{8}{n(n-1)s_n^3} \sum_{i \neq j} \left(\mathbb{E}|X_{i\pi(i)}|^3 + \mathbb{E}|X_{j\pi(j)}|^3 + \mathbb{E}|X_{i\pi(j)}|^3 + \mathbb{E}|X_{j\pi(i)}|^3 \right) \\
&= \frac{16}{n(n-1)s_n^3} \sum_{i \neq j} \left(\mathbb{E}|X_{i\pi(i)}|^3 + \mathbb{E}|X_{i\pi(j)}|^3 \right) \\
&= \frac{32}{n^2 s_n^3} \sum_{i,j=1}^{n} \mathbb{E}|X_{ij}|^3.
\end{aligned}$$

Hence,

$$\epsilon_1 \leq \frac{4\|g\|_{M^1}}{3ns_n^3} \sum_{i,j=1}^{n} \mathbb{E}|X_{ij}|^3. \tag{5.6}$$

Furthermore, by Proposition 3.2:

$$\begin{aligned}
\epsilon_3 = \left| \frac{1}{ns_n} \sum_{i,j=1}^{n} \mathbb{E}Df(\mathbf{Y}_n) \left[X_{i,\pi(j)} \mathbb{1}_{[i/n,1]} \right] \right| &= \left| \frac{1}{ns_n} \sum_{i,j=1}^{n} \mathbb{E}Df(\mathbf{Y}_n) \left[X_{i,j} \mathbb{1}_{[i/n,1]} \right] \right| \\
&\leq \|g\|_{M^1} \frac{1}{ns_n} \mathbb{E} \left\| \sum_{i,j=1}^{n} X_{i,j} \mathbb{1}_{[i/n,1]} \right\| \\
&\leq \frac{2\|g\|_{M^1}}{ns_n} \sqrt{\mathbb{E} \left| \sum_{i,j=1}^{n} X_{i,j} \right|^2} \\
&\leq \frac{2\|g\|_{M^1}}{ns_n} \sqrt{\sum_{i,j=1}^{n} \sigma_{i,j}^2} \\
&\leq \frac{2\|g\|_{M^1}}{\sqrt{n}},
\end{aligned} \tag{5.7}$$

where we have used Doob's L^2 inequality in the second inequality and (5.1) in the last one.

Step 3. Now define a new permutation π_{ijkl} coupled with π such that:

$$\mathcal{L}(\pi_{ijkl}) = \mathcal{L}(\pi|\pi(i) = k, \pi(j) = l),$$

where $\mathcal{L}(\cdot)$ denotes the law. As noted in [CF15], we can construct it in the following way. For τ_{ij} denoting the transposition of i, j:

$$\pi_{ijkl} = \begin{cases} \pi, & \text{if } l = \pi(j), k = \pi(i) \\ \pi \cdot \tau_{\pi^{-1}(k),i}, & \text{if } l = \pi(j), k \neq \pi(i) \\ \pi \cdot \tau_{\pi^{-1}(l),j}, & \text{if } l \neq \pi(j), k = \pi(i) \\ \pi \cdot \tau_{\pi^{-1}(l),i} \cdot \tau_{\pi^{-1}(k),j} \cdot \tau_{ij}, & \text{if } l \neq \pi(j), k \neq \pi(i). \end{cases}$$

We also let

$$\mathbf{Y}_{n,ijkl} = \frac{1}{s_n} \sum_{i'=1}^{n} X_{i'\pi_{ijkl}(i')} \mathbb{1}_{[i'/n,1]}.$$

Then $\mathcal{L}(\mathbf{Y}_{n,ijkl}) = \mathcal{L}(\mathbf{Y}_n|\pi(i) = k, \pi(j) = l)$ (recalling that $\mathcal{L}(\cdot)$ denotes the law). Also, for each choice of $i \neq j$, $k \neq l$ let $\mathbb{X}^{ijkl} := \left\{X_{i'j'}^{ijkl} : i', j' \in [n]\right\}$ be the same as $\mathbb{X} := \{X_{ij}; i, j \in [n]\}$ except that $\{X_{ik}, X_{il}, X_{jk}, X_{jl}\}$ has been replaced by an independent copy $\{X'_{ik}, X'_{il}, X'_{jk}, X'_{jl}\}$. Then let

$$\mathbf{Y}_n^{ijkl} = \frac{1}{s_n} \sum_{i'=1}^{n} X_{i'\pi(i')}^{ijkl} \mathbb{1}_{[i'/n,1]}$$

and note that $\mathbf{Y}_n^{ijkl}$ is independent of $\{X_{ik}, X_{il}, X_{jk}, X_{jl}\}$ and $\mathcal{L}(\mathbf{Y}_n^{ijkl}) = \mathcal{L}(\mathbf{Y}_n)$ (where $\mathcal{L}$ denotes the law).

Now, by Lemma 7.1, proved in the appendix, for ϵ_2 of Theorem 4.1,

$$\begin{aligned} \epsilon_2 = &\left|\mathbb{E}D^2 f(\mathbf{Y}_n)\left[(\mathbf{Y}_n - \mathbf{Y}'_n)\Lambda_n, \mathbf{Y}_n - \mathbf{Y}'_n\right] - \mathbb{E}D^2 f(\mathbf{Y}_n)[\mathbf{D}_n, \mathbf{D}_n]\right| \\ \leq &A + B \end{aligned} \quad (5.8)$$

where

$$\begin{aligned} A = \Bigg| &\frac{1}{n(n-1)s_n^2} \sum_{\substack{1\leq i,j,k,l\leq n \\ i\neq j, k\neq l}} \mathbb{E}\Bigg\{\left[\frac{(X_{ik} - X_{il})^2}{2n} - \frac{\hat{Z}_i^2}{n-1}\right] \\ &\qquad \cdot \left(D^2 f(\mathbf{Y}_{n,ijkl}) - D^2 f\left(\mathbf{Y}_n^{ijkl}\right)\right)\left[\mathbb{1}_{[i/n,1]}\mathbb{1}_{[i/n,1]}\right]\Bigg\} \\ &+ \frac{1}{n(n-1)s_n^2} \sum_{\substack{1\leq i,j,k,l\leq n \\ i\neq j, k\neq l}} \mathbb{E}\left\{\left[\frac{(X_{ik} - X_{il})(X_{jl} - X_{jk})}{2n} - \hat{Z}_i\hat{Z}_j\right]\right. \end{aligned}$$

$$\left.\left.\cdot\left(D^2f(\mathbf{Y}_{n,ijkl})-D^2f\left(\mathbf{Y}_n^{ijkl}\right)\right)\left[\mathbb{1}_{[i/n,1]},\mathbb{1}_{[j/n,1]}\right]\right\}\right|, \tag{5.9}$$

$$\begin{aligned}B=&\left|\frac{1}{n(n-1)s_n^2}\sum_{\substack{1\leq i,j,k,l\leq n\\ i\neq j,k\neq l}}\mathbb{E}\left\{\left[\frac{(X_{ik}-X_{il})^2}{2n}-\frac{\hat{Z}_i^2}{n-1}\right]D^2f\left(\mathbf{Y}_n^{ijkl}\right)\left[\mathbb{1}_{[i/n,1]}\mathbb{1}_{[i/n,1]}\right]\right\}\right.\\&+\frac{1}{n(n-1)s_n^2}\sum_{\substack{1\leq i,j,k,l\leq n\\ i\neq j,k\neq l}}\mathbb{E}\left\{\left[\frac{(X_{ik}-X_{il})(X_{jl}-X_{jk})}{2n}-\hat{Z}_i\hat{Z}_j\right]\right.\\&\left.\left.\cdot D^2f\left(\mathbf{Y}_n^{ijkl}\right)\left[\mathbb{1}_{[i/n,1]},\mathbb{1}_{[j/n,1]}\right]\right\}\right|.\end{aligned}$$

Recalling that $\mathbf{Y}_n^{ijkl}$ is independent of $\{X_{ik},X_{il},X_{jk},X_{jl}\}$ and $\mathcal{L}(\mathbf{Y}_n^{ijkl})=\mathcal{L}(\mathbf{Y}_n)$,

$$\begin{aligned}B=&\left|\frac{1}{n(n-1)s_n^2}\sum_{\substack{1\leq i,j,k,l\leq n\\ i\neq j,k\neq l}}\mathbb{E}\left[\frac{(X_{ik}-X_{il})^2}{2n}-\frac{\hat{Z}_i^2}{n-1}\right]\mathbb{E}\left\{D^2f\left(\mathbf{Y}_n\right)\left[\mathbb{1}_{[i/n,1]}\mathbb{1}_{[i/n,1]}\right]\right\}\right.\\&+\frac{1}{n(n-1)s_n^2}\sum_{\substack{1\leq i,j,k,l\leq n\\ i\neq j,k\neq l}}\mathbb{E}\left[\frac{(X_{ik}-X_{il})(X_{jl}-X_{jk})}{2n}-\hat{Z}_i\hat{Z}_j\right]\\&\left.\cdot\mathbb{E}\left\{D^2f\left(\mathbf{Y}_n\right)\left[\mathbb{1}_{[i/n,1]},\mathbb{1}_{[j/n,1]}\right]\right\}\right|\\&\leq\frac{\|g\|_{M^1}}{n^2(n-1)s_n^2}\sum_{1\leq i\neq j\leq n}\sum_{r=1}^n\sigma_{ir}^2\\&=\frac{\|g\|_{M^1}}{n^2s_n^2}\sum_{i,j=1}^n\sigma_{i,j}^2,\end{aligned} \tag{5.10}$$

where the inequality follows by (5.4), (5.5) and Proposition 3.2. Furthermore, by Lemma 7.2, proved in the appendix,

$$\begin{aligned}A\leq&\frac{\|g\|_{M^1}}{n^3(n-1)s_n^3}\sum_{1\leq i,j,k,l,u\leq n}\Big\{\mathbb{E}|X_{ik}|^3+5\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|^2+7\mathbb{E}|X_{ik}|^2\mathbb{E}|X_{jl}|\\&+5\mathbb{E}|X_{ik}|^2\mathbb{E}|X_{jk}|+16\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|\mathbb{E}|X_{jl}|+2\mathbb{E}|X_{iu}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|\\&+4\mathbb{E}|X_{iu}|\mathbb{E}|X_{il}|\mathbb{E}|X_{jk}|+6\mathbb{E}|X_{uk}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{jl}|+2\mathbb{E}|X_{uk}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{jk}|\end{aligned}$$

$$+\frac{1}{n}\left(2\mathbb{E}\left|X_{ik}\right|+2\mathbb{E}\left|X_{j,l}\right|+2\mathbb{E}|X_{uk}|+2\mathbb{E}|X_{ul}|\right)\cdot\sum_{r=1}^{n}\left(\mathbb{E}|X_{ir}|^2+|c_{ir}c_{jr}|\right)\Bigg\}. \tag{5.11}$$

We now use (5.6),(5.7),(5.8),(5.10),(5.11) to obtain the assertion. □

5.5. Convergence to a continuous Gaussian process

Theorem 5.5. *Let $\mathbb{X}$ and $\mathbf{Y}_n$ be as defined in Subsection 5.1 and suppose that for all $u,t\in[0,1]$:*

$$\frac{1}{s_n^2(n-1)}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{j=1}^{\lfloor nu\rfloor}\sum_{k=1}^{n}\mathbb{E}X_{ik}X_{jk}\left(\delta_{i,j}-\frac{1}{n}\right)\xrightarrow{n\to\infty}\sigma(u,t) \tag{5.12}$$

and

$$\frac{1}{s_n^2}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{j=1}^{\lfloor nu\rfloor}\sum_{l=1}^{n}\mathbb{E}X_{il}X_{jl}\xrightarrow{n\to\infty}\sigma^{(2)}(u,t) \tag{5.13}$$

pointwise for some functions $\sigma,\sigma^{(2)}:[0,1]^2\to\mathbb{R}_+$. Suppose furthermore that:

$$\sup_{n\in\mathbb{N}}\frac{1}{n^2s_n^4}\sum_{l=1}^{n}\sum_{i=1}^{n}Var\left[X_{il}^2\right]<\infty. \tag{5.14}$$

and:

$$\frac{1}{s_n^2(n-1)}\sum_{i=1}^{\lfloor nt\rfloor}\left(\sum_{l=1}^{n}X_{il}''Z_{il}\right)^2\xrightarrow{P}c(t) \tag{5.15}$$

pointwise for some function $c:[0,1]\to\mathbb{R}_+$ and:

$$\lim_{n\to\infty}\frac{1}{s_n\sqrt{n-1}}\mathbb{E}\left[\sup_{i=1,\cdots,n}|X_{il}''Z_{il}|\right]=0. \tag{5.16}$$

Then $(\mathbf{Y}_n(t),t\in[0,1])$ converges weakly in the uniform topology to a continuous Gaussian process $(\mathbf{Z}(t),t\in[0,1])$ with the covariance function σ.

Remark 5.6. *Assumption (5.13) could also say that*

$$\frac{1}{s_n^2}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{j=1}^{\lfloor nu\rfloor}\sum_{l=1}^{n}\mathbb{E}X_{il}X_{jl}$$

simply converges pointwise rather than giving the limit a name. However, we will use $\sigma^{(2)}$ in the proof so it is convenient to use it in the formulation of the Theorem as well.

Remark 5.7. *Assumption (5.15) is necessary for the limiting process in Theorem 5.5 to be continuous. It essentially corresponds to the the assumption that the quadratic variation of the following process*

$$\mathbf{D}_n^{(1)}(t) = \frac{1}{s_n\sqrt{n-1}}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{l=1}^{n} X''_{il}Z_{il}$$

converges to the function c pointwise in probability, which then implies the weak convergence of the process $\mathbf{D}_n^{(1)}$ to a continuous process. While it is relatively easy to show that $\mathbf{D}_n^{(2)} = \mathbf{D}_n - \mathbf{D}_n^{(1)}$ converges to a continuous limit, we had to explicitly add this assumption to ensure that $\mathbf{D}_n$ does as well.

The proof of Theorem 5.5 will be similar to the proof of [BJ09, Theorem 3.3]. The pre-limiting approximand $\mathbf{D}_n$, defined in Subsection 5.3, will be expressed as a sum of two parts. In **Steps 1** and **2** we prove that each of those parts is C-tight (i.e. they are tight and for each of them any convergent subsequence converges to a process with continuous sample paths). In **Step 3** we show that the assumptions of Theorem 5.5 trivially imply the convergence of the covariance function of $\mathbf{D}_n$, which together with C-tightness implies the convergence of $\mathbf{D}_n$ to a continuous process. Theorem 6.2 will then be combined with Proposition 2.3 to show convergence of $\mathbf{Y}_n$ to the same limiting process. Finally, the combinatorial central limit theorem for random arrays, proved in [HC78] and analysed in [CF15], will imply that $\mathbf{Z}$ is Gaussian.

Proof of Theorem 5.5. We will use the notation of Subsections 5.1 and 5.3.

Step 1. Note that $\mathbf{D}_n = \mathbf{D}_n^{(1)} + \mathbf{D}_n^{(2)}$, where:

$$\mathbf{D}_n^{(1)}(t) = \frac{1}{s_n\sqrt{n-1}}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{l=1}^{n} X''_{il}Z_{il}, \quad \mathbf{D}_n^{(2)}(t) = \frac{1}{s_n\sqrt{n-1}}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{l=1}^{n} X''_{il}\bar{Z}_l$$

for $\bar{Z}_l = \frac{1}{n}\sum_{j=1}^n Z_{jl}$.

Now, note that, by (5.15):

$$\left\langle \mathbf{D}_n^{(1)} \right\rangle_t \xrightarrow{P} c(t)$$

pointwise, where $\langle\cdot\rangle$ denotes quadratic variation. Therefore, by [EK86, Chapter 7, Theorem 1.4] and using (5.16), we obtain that $\mathbf{D}_n^{(1)}$ converges weakly in the Skorokhod topology on $D[0,1]$ to a continuous Gaussian process with independent increments.

We now note that the Skorokhod space equipped with the metric (topologically equivalent to the Skorokhod metric) with respect to which it is complete is also universally measurable by the discussion at the beginning of [Dud02, Chapter 11.5]. Since it is also separable and $\mathbf{D}_n^{(1)} \Rightarrow \mathbf{Z}_1$, for some continuous process $\mathbf{Z}_1$, in the Skorokhod topology, [Dud02, Theorem 11.5.3] implies that $(\mathbf{D}_n^{(1)})_{n\geq 1}$ is C-tight.

Step 2. Also, note that for $u > t$ s.t. $\lfloor nu \rfloor \geq \lfloor nt \rfloor + 1$,

$$\begin{aligned}\mathbb{E}\left[\left|\mathbf{D}_n^{(2)}(u) - \mathbf{D}_n^{(2)}(t)\right|^2 \middle| X''_{il}, i,l \in [n]\right] =& \frac{1}{n(n-1)s_n^2}\sum_{l=1}^{n}\left(\sum_{i=\lfloor nt \rfloor+1}^{\lfloor nu \rfloor} X_{il}\right)^2 \\ \leq& \frac{\lfloor nu \rfloor - \lfloor nt \rfloor}{n(n-1)s_n^2}\sum_{l=1}^{n}\sum_{i=\lfloor nt \rfloor+1}^{\lfloor nu \rfloor} X_{il}^2\end{aligned}$$

and

$$\mathbb{E}\left[\left|\mathbf{D}_n^{(2)}(u) - \mathbf{D}_n^{(2)}(t)\right|^2 \middle| X''_{il}, i,l \in [n]\right] = 0, \quad \text{for } u > t \text{ s.t. } \lfloor nu \rfloor = \lfloor nt \rfloor.$$

Since $\left(\mathbf{D}_n^{(2)} | X''_{il}, i,l \in [n]\right)$ is Gaussian for u, such that $\lfloor nu \rfloor \geq \lfloor nt \rfloor + 1$,

$$\begin{aligned}&\mathbb{E}\left|\mathbf{D}_n^{(2)}(u) - \mathbf{D}_n^{(2)}(t)\right|^4 \\ =&3\mathbb{E}\left\{\left(\mathbb{E}\left[\left|\mathbf{D}_n^{(2)}(u) - \mathbf{D}_n^{(2)}(t)\right|^2 \middle| X''_{il}, i,l \in [n]\right]\right)^2\right\} \\ \leq&3\left(\frac{\lfloor nu \rfloor - \lfloor nt \rfloor}{n(n-1)s_n^2}\right)^2 \mathbb{E}\left(\sum_{l=1}^{n}\sum_{i=\lfloor nt \rfloor+1}^{\lfloor nu \rfloor} X_{il}^2\right)^2 \\ =&3\left(\frac{\lfloor nu \rfloor - \lfloor nt \rfloor}{n(n-1)s_n^2}\right)^2\left[\left(\sum_{l=1}^{n}\sum_{i=\lfloor nt \rfloor+1}^{\lfloor nu \rfloor}\mathbb{E}X_{il}^2\right)^2 + \sum_{l=1}^{n}\sum_{i=\lfloor nt \rfloor+1}^{\lfloor nu \rfloor}\left(\mathbb{E}X_{il}^4 - \left(\mathbb{E}X_{il}^2\right)^2\right)\right] \\ \leq&C\left(\frac{\lfloor nu \rfloor - \lfloor nt \rfloor}{(n-1)}\right)^2 \end{aligned} \tag{5.17}$$

for some constant C, by (5.14). Now, note that:

$$\operatorname{Cov}\left(\mathbf{D}_n^{(2)}(t), \mathbf{D}_n^{(2)}(u)\right) = \frac{1}{s_n^2}\sum_{i=1}^{\lfloor nt \rfloor}\sum_{j=1}^{\lfloor nu \rfloor}\sum_{l=1}^{n}\mathbb{E}X_{il}X_{jl} \xrightarrow{n\to\infty} \sigma^{(2)}(t,u),$$

by (5.13). Consider a mean zero Gaussian process $\mathbf{Z}_2$ with covariance function $\mathbb{E}\mathbf{Z}_2(t)\mathbf{Z}_2(u) = \sigma^{(2)}(t,u)$. The finite dimensional distributions of $\mathbf{D}_n^{(2)}$ converge to those of $\mathbf{Z}_2$. We can now construct $\mathbf{D}_n^{(2)}$ and $\mathbf{Z}_2$ on the same probability space and use Skorokhod's representation theorem, Fatou's lemma and (7.4) to conclude that:

$$\mathbb{E}\left(|\mathbf{Z}_2(u) - \mathbf{Z}_2(t)|^4\right) \leq \lim_{n\to\infty}\mathbb{E}\left(\left|\mathbf{D}_n^{(2)}(u) - \mathbf{D}_n^{(2)}(t)\right|^4\right) \leq C(u-t)^2.$$

By [Bil68, Theorem 12.4], we can assume that $\mathbf{Z}_2 \in C[0,1]$. Now, note that for $0 \le t \le v \le u \le 1$:

$$\begin{aligned}
&\mathbb{E}\left|\mathbf{D}_n^{(2)}(v) - \mathbf{D}_n^{(2)}(t)\right|^2 \left|\mathbf{D}_n^{(2)}(v) - \mathbf{D}_n^{(2)}(u)\right|^2 \\
\le&\sqrt{\mathbb{E}\left|\mathbf{D}_n^{(2)}(v) - \mathbf{D}_n^{(2)}(t)\right|^4 \mathbb{E}\left|\mathbf{D}_n^{(2)}(v) - \mathbf{D}_n^{(2)}(u)\right|^4} \\
\overset{(7.4)}{\le}& C\frac{(\lfloor nv\rfloor - \lfloor nt\rfloor)(\lfloor nu\rfloor - \lfloor nv\rfloor)}{(n-1)^2} \\
\le&\bar{C}(u-t)^2;
\end{aligned}$$

for some constant $\bar{C}$. Therefore, by [Bil68, Theorem 15.6], $\mathbf{D}_n^{(2)} \Rightarrow \mathbf{Z}_2$ in the Skorokhod and uniform topologies and so, by [Dud02, Theorem 11.5.3], $\mathbf{D}_n^{(2)}$ is C-tight.

Step 3. Since both $\mathbf{D}_n^{(1)}$ and $\mathbf{D}_n^{(2)}$ are C-tight, so is their difference $\mathbf{D}_n$. Now:

$$\begin{aligned}
\mathrm{Cov}(\mathbf{D}_n(t), \mathbf{D}_n(u)) =&\frac{1}{s_n^2(n-1)}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{j=1}^{\lfloor nu\rfloor}\sum_{k,l=1}^{n}\mathbb{E}\left\{X_{ik}X_{jl}\left(Z_{ik}-\bar{Z}_k\right)\left(Z_{jl}-\bar{Z}_l\right)\right\} \\
=&\frac{1}{s_n^2(n-1)}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{j=1}^{\lfloor nu\rfloor}\sum_{k=1}^{n}\mathbb{E}\left\{X_{ik}X_{jk}\left(Z_{ik}-\bar{Z}_k\right)\left(Z_{jk}-\bar{Z}_k\right)\right\} \\
=&\frac{1}{s_n^2(n-1)}\sum_{i=1}^{\lfloor nt\rfloor}\sum_{j=1}^{\lfloor nu\rfloor}\sum_{k=1}^{n}\mathbb{E}X_{ik}X_{jk}\left(\delta_{i,j}-\frac{1}{n}\right)\xrightarrow{n\to\infty}\sigma(u,t),
\end{aligned}$$

by (5.12) and we obtain that $\mathbf{D}_n$ converges to a random element $\mathbf{Z} \in C[0,1]$ with covariance function σ in distribution with respect to the uniform and Skorokhod topologies.

Proposition 2.3 and Theorem 5.1 therefore imply that $(\mathbf{Y}_n(t), t \in [0,1])$ converges weakly to $(\mathbf{Z}(t), t \in [0,1])$ in the uniform topology. Using, for example, [CF15, Theorem 1.1], we conclude that $\mathbf{Z}$ is a Gaussian process. □

6. Edge and two-star counts in Bernoulli random graphs

In this section we consider a two-dimensional process whose first coordinate is a properly rescaled number of edges and the second one is a rescaled number of two-stars (i.e. subgraphs which are trees with one internal node and 2 leaves) in a Bernoulli random graph with a fixed edge probability and $\lfloor nt\rfloor$ edges for $t \in [0,1]$. A similar setup has been considered in [RR10], where the authors established a bound on the distance between a three-dimensional vector consisting of a rescaled number of edges, a rescaled number of two-stars and a rescaled number of triangles in a $G(n,p)$ graph and a three-dimensional Gaussian vector. We first compare our process to a two-dimensional Gaussian pre-limiting Gaussian processes with paths in $D([0,1])$ and bound the distance between the two in Theorem 6.2. Then, in Theorem 6.4, we bound the distance of our process from a continuous two-dimensional Gaussian process.

It is worth noting that the analysis of this section could easily be extended to one of a three-dimensional process whose coordinates represent the number of edges, the number of triangles and the number of two-stars in a $G(\lfloor nt\rfloor, p)$ graph. The only reason we do not do it here is that it would require some more involved algebraic computations and would make this section rather lengthy.

6.1. Introduction

Let us consider a Bernoulli random graph $G(n,p)$ on n vertices with edge probabilities p.

Let $I_{i,j}=I_{j,i}$ be the Bernoulli(p)-indicator that edge (i,j) is present in this graph. These indicators, for $(i,j)\in\{1,\cdots,n\}^2$ are independent. We will look at a process representing at each $t\in[0,1]$ the re-scaled total number of edges in the graph formed out of the given Bernoulli random graph by considering only its first $\lfloor nt\rfloor$ vertices and the edges between them:

$$\mathbf{T}_n(t)=\frac{\lfloor nt\rfloor-2}{2n^2}\sum_{i,j=1}^{\lfloor nt\rfloor}I_{i,j}=\frac{\lfloor nt\rfloor-2}{n^2}\sum_{1\leq i<j\leq\lfloor nt\rfloor}I_{i,j},$$

and at a process representing a re-scaled statistic related to the number of two-stars in the same graph:

$$\mathbf{V}_n(t)=\frac{1}{2n^2}\sum_{\substack{1\leq i,j,k\leq\lfloor nt\rfloor\\ i,j,k\text{ distinct}}}I_{ij}I_{jk}=\frac{1}{n^2}\sum_{1\leq i<j<k\leq\lfloor nt\rfloor}\left(I_{i,j}I_{j,k}+I_{i,j}I_{i,k}+I_{j,k}I_{i,k}\right).$$

Let $\mathbf{Y}_n(t)=(\mathbf{T}_n(t)-\mathbb{E}\mathbf{T}_n(t),\mathbf{V}_n(t)-\mathbb{E}\mathbf{V}_n(t))$ for $t\in[0,1]$.

Remark 6.1. *Note that, for all* $t\in[0,1]$, $\mathbb{E}\mathbf{T}_n(t)=\frac{\lfloor nt\rfloor-2}{n^2}\binom{\lfloor nt\rfloor}{2}p$ *and* $\mathbb{E}\mathbf{V}_n(t)=\frac{3}{n^2}\binom{\lfloor nt\rfloor}{3}p^2$. *Furthermore, note that, by an argument similar to that of [RR10, Section 5], the covariance matrix of* $(\mathbf{T}_n(t)-\mathbb{E}\mathbf{T}_n(t),\mathbf{V}_n(t)-\mathbb{E}\mathbf{V}_n(t))$ *is given by*

$$3\frac{(\lfloor nt\rfloor-2)\binom{\lfloor nt\rfloor}{3}}{n^4}p(1-p)\begin{pmatrix}1 & 2p\\ 2p & 4p^2\end{pmatrix}.$$

Hence, the scaling ensures that the covariances are of the same order in n.

6.2. Exchangeable pair setup

We now construct an exchangeable pair, as in [RR10], by picking (I,J) according to $\mathbb{P}[I=i,J=j]=\frac{1}{\binom{n}{2}}$ for $1\leq i<j\leq n$. If $I=i$, $J=j$, we replace $I_{i,j}=I_{j,i}$

by an independent copy $I'_{i,j} = I'_{j,i}$ and put:

$$\mathbf{T}'_n(t) = \mathbf{T}_n(t) - \frac{\lfloor nt \rfloor - 2}{n^2}\left(I_{I,J} - I'_{I,J}\right)\mathbb{1}_{[I/n,1]\cap[J/n,1]}(t)$$
$$\mathbf{V}'_n(t) = \mathbf{V}_n(t) - \frac{1}{n^2}\sum_{k:k\neq I,J}\left(I_{I,J} - I'_{I,J}\right)(I_{J,k} + I_{I,k})\,\mathbb{1}_{[I/n,1]\cap[J/n,1]\cap[k/n,1]}(t).$$

We also let $\mathbf{Y}'_n(t) = (\mathbf{T}'_n(t) - \mathbb{E}\mathbf{T}_n(t), \mathbf{V}'_n(t) - \mathbb{E}\mathbf{V}_n(t))$ and note that, for $\mathbf{Y}_n = (\mathbf{Y}_n(t), t \in [0,1])$ and $\mathbf{Y}'_n = (\mathbf{Y}'_n(t), t \in [0,1])$, $(\mathbf{Y}_n, \mathbf{Y}'_n)$ forms an exchangeable pair. Let $e_1 = (1,0)$, $e_2 = (0,1)$. We note that, for any $m = 1,2$ and for any $f \in M$, as defined in Section 2,

$$\begin{aligned}
&\mathbb{E}^{\mathbf{Y}_n}\left\{Df(\mathbf{Y}_n)\left[(\mathbf{T}'_n - \mathbf{T}_n)\,e_m\right]\right\}\\
=&\mathbb{E}^{\mathbf{Y}_n}\left\{Df(\mathbf{Y}_n)\left[\frac{\lfloor n\cdot\rfloor - 2}{n^2}\left(I'_{I,J} - I_{I,J}\right)\mathbb{1}_{[I/n,1]\cap[J/n,1]}e_m\right]\right\}\\
=&\frac{2}{n^3(n-1)}\sum_{i<j}\mathbb{E}^{\mathbf{Y}_n}\left\{Df(\mathbf{Y}_n)\left[(\lfloor n\cdot\rfloor - 2)\left(I'_{i,j} - I_{i,j}\right)\mathbb{1}_{[i/n,1]\cap[j/n,1]}e_m\right]|I=i, J=j\right\}\\
=&-\frac{1}{\binom{n}{2}}Df(\mathbf{Y}_n)[\mathbf{T}_n e_m] + \frac{2}{n^3(n-1)}p\sum_{i<j}Df(\mathbf{Y}_n)\left[(\lfloor n\cdot\rfloor - 2)\mathbb{1}_{[i/n,1]\cap[j/n,1]}e_m\right]\\
=&-\frac{1}{\binom{n}{2}}Df(\mathbf{Y}_n)[(\mathbf{T}_n(\cdot) - \mathbb{E}\mathbf{T}_n(\cdot))\,e_m].
\end{aligned}$$

Also:

$$\begin{aligned}
&\mathbb{E}^{\mathbf{Y}_n}Df(\mathbf{Y}_n)[(\mathbf{V}_n - \mathbf{V}'_n)e_m]\\
=&\frac{1}{n^2\binom{n}{2}}\sum_{i<j}\mathbb{E}^{\mathbf{Y}_n}\left\{\sum_{k:k\neq i,j}Df(\mathbf{Y}_n)\left[\left(I_{i,j} - I'_{i,j}\right)(I_{j,k} + I_{i,k})\right.\right.\\
&\left.\left.\cdot\ \mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}e_m\right]\Big|\, I=i, J=j\right\}\\
=&\frac{2}{\binom{n}{2}}Df(\mathbf{Y}_n)[\mathbf{V}_n e_m]\\
&-\frac{p}{n^2\binom{n}{2}}\sum_{i<j}\sum_{k:k\neq i,j}\mathbb{E}^{\mathbf{Y}_n}Df(\mathbf{Y}_n)\left[(I_{j,k} + I_{i,k})\,\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}e_m\right]\\
=&\frac{2}{\binom{n}{2}}Df(\mathbf{Y}_n)[\mathbf{V}_n e_m] - \frac{p}{n^2\binom{n}{2}}\sum_{\substack{1\leq i,j,k\leq n\\ i,j,k\text{ distinct}}}\mathbb{E}^{\mathbf{Y}_n}Df(\mathbf{Y}_n)\left[I_{i,j}\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}e_m\right]\\
=&\frac{2}{\binom{n}{2}}Df(\mathbf{Y}_n)[(\mathbf{V}_n - \mathbb{E}\mathbf{V}_n(\cdot))\,e_m]\\
&-\frac{p}{n^2\binom{n}{2}}\sum_{\substack{1\leq i,j,k\leq n\\ i,j,k\text{ distinct}}}\mathbb{E}^{\mathbf{Y}_n}Df(\mathbf{Y}_n)\left[(I_{i,j} - p)\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}e_m\right]
\end{aligned}$$

$$=\frac{2}{\binom{n}{2}}Df(\mathbf{Y}_n)[(\mathbf{V}_n-\mathbb{E}\mathbf{V}_n(\cdot))\,e_m]$$
$$-\frac{2p}{\binom{n}{2}}Df(\mathbf{Y}_n)\left[\frac{1}{\lfloor n\cdot\rfloor-2}\left(\mathbf{T}_n-\mathbb{E}\mathbf{T}_n(\cdot)\right)e_m\left(\sum_{k=1}^{n}\mathbb{1}_{[k/n,1]}-2\right)\right]$$
$$=\frac{2}{\binom{n}{2}}Df(\mathbf{Y}_n)[(\mathbf{V}_n-\mathbb{E}\mathbf{V}_n(\cdot))\,e_m]-\frac{2p}{\binom{n}{2}}Df(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbb{E}\mathbf{T}_n(\cdot))e_m\right].$$

Therefore, for any $m=1,2$:

A) $Df(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbb{E}\mathbf{T}_n)\,e_m\right]=\dfrac{n(n-1)}{2}\mathbb{E}^{\mathbf{Y}_n}\left\{Df(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbf{T}_n')e_m\right]\right\}$

B) $Df(\mathbf{Y}_n)\left[(\mathbf{V}_n-\mathbb{E}\mathbf{V}_n)\,e_m\right]$

$$=\frac{n(n-1)}{4}\mathbb{E}^{\mathbf{Y}_n}\left\{Df(\mathbf{Y}_n)\left[(\mathbf{V}_n-\mathbf{V}_n')e_m\right]+pDf(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbb{E}\mathbf{T}_n)\,e_m\right]\right\}$$
$$=\frac{n(n-1)}{4}\mathbb{E}^{\mathbf{Y}_n}\left\{Df(\mathbf{Y}_n)\left[(2p(\mathbf{T}_n-\mathbf{T}_n')+\mathbf{V}_n-\mathbf{V}_n')\,e_m\right]\right\}$$

and so:

$$Df(\mathbf{Y}_n)[\mathbf{Y}_n]=2\mathbb{E}^{\mathbf{Y}_n}Df(\mathbf{Y}_n)\left[(\mathbf{Y}_n-\mathbf{Y}_n')\Lambda_n\right],$$

where:

$$\Lambda_n=\frac{n(n-1)}{8}\begin{pmatrix}2 & 2p\\ 0 & 1\end{pmatrix}. \tag{6.1}$$

Therefore, condition (4.1) is satisfied with Λ_n of (6.1) and $R_f=0$.

6.3. A pre-limiting process

Let $\mathbf{D}_n=(\mathbf{D}_n^{(1)},\mathbf{D}_n^{(2)})$, where $\mathbf{D}_n^{(2)}=\mathbf{D}_n^{(2,1)}+\mathbf{D}_n^{(2,2)}$, be defined in the following way:

$$\mathbf{D}_n^{(1)}(t)=(\lfloor nt\rfloor-2)\sum_{i,j=1}^{\lfloor nt\rfloor}Z_{i,j}^{(1)},\quad t\in[0,1]$$
$$\mathbf{D}_n^{(2,1)}(t)=(\lfloor nt\rfloor-2)\sum_{i,j=1}^{\lfloor nt\rfloor}Z_{i,j}^{(2,1)},\quad t\in[0,1]$$
$$\mathbf{D}_n^{(2,2)}(t)=\sum_{i,j,k=1}^{\lfloor nt\rfloor}Z_{i,j,k}^{(2,2)},\quad t\in[0,1]$$

where $Z_{i,i}^{(1)}=0$ for all i, $Z_{i,i}^{(1,2)}=0$ for all i and $Z_{i,j,k}^{(2,2)}=0$ if $i=j$ or $i=k$ of $j=k$. Furthermore, assume that the collection $\{Z_{i,j}^{(1)}:i,j\in[n],i\neq j\}\cup\{Z_{i,j}^{(1,2)}:i,j\in[n],i\neq j\}\cup\{Z_{i,j,k}^{(2,2)}:i,j,k\in[n],i\neq j\neq k\neq i\}$ is jointly centred Gaussian with the following covariance structure:

$$\mathbb{E}Z^{(1)}_{ij}Z^{(1)}_{kl}=\begin{cases}\frac{p(1-p)}{2n^4}, & i=k,j=l,i\neq j\\ 0, & \text{otherwise},\end{cases}$$

$$\mathbb{E}Z^{(1)}_{i,j}Z^{(2,1)}_{k,l}=\begin{cases}\frac{p^2(1-p)}{4n^4}, & i=k,j=l,i\neq j\\ 0, & \text{otherwise},\end{cases}$$

$$\mathbb{E}Z^{(2,2)}_{i,j,k}Z^{(1)}_{l,m}=\begin{cases}\frac{3p^2(1-p)}{4n^4}, & i=l,j=m,i\neq j\neq k\neq i\\ 0, & \text{otherwise},\end{cases}$$

$$\mathbb{E}Z^{(2,2)}_{i,j,k}Z^{(2,1)}_{l,m}=\begin{cases}\frac{p^3(1-p)}{2n^4}, & i=l,j=m,i\neq j\neq k\neq i\\ 0, & \text{otherwise},\end{cases}$$

$$\mathbb{E}Z^{(2,2)}_{i,j,k}Z^{(2,2)}_{r,s,t}=\begin{cases}\frac{p^2(1-p^2)}{2n^4}, & i=r,j=s,k=t,i\neq j\neq k\neq i\\ \frac{p^3(1-p)}{n^4}, & i=r,j=s,k\neq t,i\neq j\neq k\neq i,i\neq j\neq t\neq i\\ 0, & \text{otherwise},\end{cases}$$

$$\mathbb{E}Z^{(2,1)}_{i,j}Z^{(2,1)}_{k,l}=\begin{cases}\frac{1}{n^5}, & i=k,j=l,i\neq j\\ 0, & \text{otherwise}.\end{cases}$$

It will become clear in Remark 6.3 why we have chosen this covariance structure.

6.4. Distance from the pre-limiting process

We first give a theorem providing a bound on the distance between $\mathbf{Y}_n$ and the pre-limiting piecewise constant Gaussian process.

Theorem 6.2. *Let $\mathbf{Y}_n$ be defined as in Section 6.1 and $\mathbf{D}_n$ be defined as in Section 6.3. Then, for any $g\in M^2$, as defined by (2.3),*

$$|\mathbb{E}g(\mathbf{Y}_n)-\mathbb{E}g(\mathbf{D}_n)|\leq 12\|g\|_{M^2}n^{-1}.$$

In **Step 1** of the proof, which is based on Theorem 4.1, we estimate term ϵ_1 thereof. It involves bounding $\|\Lambda_n\|_2$ of (6.1) and the third moment of $\|\mathbf{Y}_n-\mathbf{Y}_n'\|$ for $\mathbf{Y}_n'$ constructed in Section 6.2. In **Step 2** we treat ϵ_2, which requires involved calculations, based on Stein's method, which are, to a large extent, postponed to the appendix. Term ϵ_3 is equal to zero as R_f of Section 6.2 is equal to zero.

Proof of Theorem 6.2. We adopt the notation of sections 6.1, 6.2, 6.3. We will apply Theorem 4.1.

Step 1. First note that, for ϵ_1 in Theorem 4.1,

$$|(\mathbf{Y}_n-\mathbf{Y}_n')\Lambda_n|\leq\|\Lambda_n\|_2|\mathbf{Y}_n-\mathbf{Y}_n'|,$$

where $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^2$ and $\|\cdot\|_2$ is the induced operator 2-norm. Furthermore, for $\|\cdot\|_F$ denoting the Frobenius norm (which, for $\Theta\in$

$\mathbb{R}^{d_1\times d_2}$ is defined by $\|\Theta\|_F=\sqrt{\sum_{i=1}^{d_1}\sum_{j=1}^{d_2}|\Theta_{i,j}|}$),

$$\|\Lambda_n\|_2\leq\|\Lambda_n\|_F=\frac{n(n-1)}{8}\sqrt{2^2+(2p)^2+0^2+1^2}\leq\frac{3n(n-1)}{8}.$$

Therefore:

$$\begin{aligned}&\mathbb{E}\|(\mathbf{Y}_n-\mathbf{Y}'_n)\Lambda_n\|\|\mathbf{Y}_n-\mathbf{Y}'_n\|^2\\ \leq&\frac{3n(n-1)}{8}\mathbb{E}\|\mathbf{Y}_n-\mathbf{Y}'_n\|^3\\ \leq&\frac{3n(n-1)}{8}\mathbb{E}\left[\frac{(n-2)^2}{n^4}\left(I_{I,J}-I'_{I,J}\right)^2+\frac{1}{n^4}\left(\sum_{k:k\neq I,J}(I_{I,J}-I'_{I,J})\left(I_{J,k}+I_{I,k}\right)\right)^2\right]^{3/2}\\ \leq&\frac{3n(n-1)}{8}\left[\frac{(n-2)^2}{n^4}+\frac{(2(n-2))^2}{n^4}\right]^{3/2}\\ \leq&\frac{5}{n},\end{aligned}\tag{6.2}$$

where the third inequality follows because $|I_{I,J}-I'_{I,J}|\leq 1$ and $|I_{J,k}+I_{I,k}|\leq 2$ for all k. Therefore,

$$\epsilon_1\leq\frac{5\|g\|_{M^2}}{6n}.$$

Step 2. For ϵ_2 in Theorem 4.1, we wish to bound:

$$\begin{aligned}&\left|\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{Y}_n-\mathbf{Y}'_n)\Lambda_n,\mathbf{Y}_n-\mathbf{Y}'_n\right]-\mathbb{E}D^2f(\mathbf{Y}_n)\left[\mathbf{D}_n,\mathbf{D}_n\right]\right|\\ =&\left|\frac{n(n-1)}{8}\mathbb{E}D^2f(\mathbf{Y}_n)\left[(2(\mathbf{T}_n-\mathbf{T}'_n),2p(\mathbf{T}_n-\mathbf{T}'_n)+(\mathbf{V}_n-\mathbf{V}'_n)),(\mathbf{T}_n-\mathbf{T}'_n,\mathbf{V}_n-\mathbf{V}'_n)\right]\right.\\ &\left.-\mathbb{E}D^2f(\mathbf{Y}_n)\left[\mathbf{D}_n,\mathbf{D}_n\right]\right|\\ \leq&S_1+S_2+S_3+S_4+S_5+S_6+S_7,\end{aligned}\tag{6.3}$$

where:

$$S_1=\left|\frac{n(n-1)}{8}\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbf{T}'_n)(2,0),(\mathbf{T}_n-\mathbf{T}'_n)(1,0)\right]-\mathbb{E}D^2f(\mathbf{Y}_n)\left[\left(\mathbf{D}_n^{(1)},0\right),\left(\mathbf{D}_n^{(1)},0\right)\right]\right|$$

$$S_2=\left|\frac{n(n-1)}{8}\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbf{T}'_n)(0,2p),(\mathbf{T}_n-\mathbf{T}'_n)(1,0)\right]-2\mathbb{E}D^2f(\mathbf{Y}_n)\left[\left(0,\mathbf{D}_n^{(2,1)}\right),\left(\mathbf{D}_n^{(1)},0\right)\right]\right|$$

$$S_3=\left|\frac{n(n-1)}{8}\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbf{T}'_n)(2,0),(\mathbf{V}_n-\mathbf{V}'_n)(0,1)\right]-\frac{4}{3}\mathbb{E}D^2f(\mathbf{Y}_n)\left[\left(\mathbf{D}_n^{(1)},0\right),\left(0,\mathbf{D}_n^{(2,2)}\right)\right]\right|$$

$$S_4=\left|\frac{n(n-1)}{8}\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbf{T}'_n)(0,2p),(\mathbf{V}_n-\mathbf{V}'_n)(0,1)\right]-2\mathbb{E}D^2f(\mathbf{Y}_n)\left[\left(0,\mathbf{D}_n^{(2,1)}\right),\left(0,\mathbf{D}_n^{(2,2)}\right)\right]\right|$$

$$S_5=\left|\frac{n(n-1)}{8}\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{V}_n-\mathbf{V}'_n)(0,1),(\mathbf{T}_n-\mathbf{T}'_n)(1,0)\right]-\frac{2}{3}\mathbb{E}D^2f(\mathbf{Y}_n)\left[\left(0,\mathbf{D}_n^{(2,2)}\right),\left(\mathbf{D}_n^{(1)},0\right)\right]\right|$$

$$S_6 = \left|\frac{n(n-1)}{8}\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{V}_n-\mathbf{V}_n')(0,1),(\mathbf{V}_n-\mathbf{V}_n')(0,1)\right]-\mathbb{E}D^2f(\mathbf{Y}_n)\left[\left(0,\mathbf{D}_n^{(2,2)}\right),\left(0,\mathbf{D}_n^{(2,2)}\right)\right]\right|$$

$$S_7 = \left|\mathbb{E}D^2f(\mathbf{Y}_n)\left[\left(0,\mathbf{D}_n^{(2,1)}\right),\left(0,\mathbf{D}_n^{(2,1)}\right)\right]\right|. \tag{6.4}$$

The following bounds are obtained in Lemma 7.3, in the appendix:

$$S_1 \leq \frac{\sqrt{5}\|g\|_{M^2}}{12n}, \quad S_2 \leq \frac{\sqrt{5}\|g\|_{M^2}}{12n}, \quad S_3 \leq \frac{\sqrt{178}\|g\|_{M^2}}{6n}, \quad S_4 \leq \frac{\sqrt{178}\|g\|_{M^2}}{6n}$$
$$S_5 \leq \frac{\sqrt{178}\|g\|_{M^2}}{12n}, \quad S_6 \leq \frac{\sqrt{612}\|g\|_{M^2}}{6n}, \quad S_7 \leq \frac{\|g\|_{M^2}}{n}. \tag{6.5}$$

Note that, by (6.3) and (6.5),

$$\begin{aligned}&\left|\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{Y}_n-\mathbf{Y}_n')\Lambda_n,\mathbf{Y}_n-\mathbf{Y}_n'\right]-\mathbb{E}D^2f(\mathbf{Y}_n)\left[\mathbf{D}_n,\mathbf{D}_n\right]\right|\\ =&\left|\frac{n(n-1)}{8}\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbf{T}_n')(2,2p)+(\mathbf{V}_n-\mathbf{V}_n')(0,1),(\mathbf{T}_n-\mathbf{T}_n')(1,0)+(\mathbf{V}_n-\mathbf{V}_n')(0,1)\right]\right.\\ \leq& 11\|g\|_{M^2}n^{-1}.\end{aligned} \tag{6.6}$$

Using Theorem 4.1 together with (6.6) and (6.2) gives the desired result. □

Remark 6.3. *The reasons for the covariance structure of $\mathbf{D}_n$ taking the particular form described in Section 6.3 become clear when we look at (6.4) and (6.5). The processes we compare are two-dimensional. The $\mathbf{D}_n^{(1)}$-part of the pre-limiting process $\mathbf{D}_n$ corresponds to the contribution of $\mathbf{T}_n-\mathbf{T}_n'$ to the first coordinate in processes $(\mathbf{Y}_n-\mathbf{Y}_n')\Lambda_n$ and $\mathbf{Y}_n-\mathbf{Y}_n'$. Similarly, $\mathbf{D}_n^{(2,1)}$ corresponds to the contribution of $\mathbf{T}_n-\mathbf{T}_n'$ to the second coordinate and $\mathbf{D}_n^{(2,2)}$ corresponds to the contribution of $\mathbf{V}_n-\mathbf{V}_n'$ to the second coordinate.*

The covariances are chosen so that at any time points $s,t\in[0,1]$, $Cov(\mathbf{D}_n(s),\mathbf{D}_n(t))$ is close to $Cov((\mathbf{Y}_n-\mathbf{Y}_n')\Lambda_n(s),(\mathbf{Y}_n-\mathbf{Y}_n')(t))$. This makes the bounds in (6.5) small. Specifically, the only contribution to

$$Cov(\mathbf{D}_n(s),\mathbf{D}_n(t))-Cov((\mathbf{Y}_n-\mathbf{Y}_n')\Lambda_n(s),(\mathbf{Y}_n-\mathbf{Y}_n')(t))$$

for $s,t\in[0,1]$ comes from the covariance of $\mathbf{D}_n^{(1)}$ and this is achieved by choosing specific values for $Cov\left(\mathbf{D}_n^{(2)}(s),\mathbf{D}_n^{(2)}(t)\right)$ and $Cov\left(\mathbf{D}_n^{(1)}(s),\mathbf{D}_n^{(2)}(t)\right)$ for $s,t\in[0,1]$.

The covariance structure of $\mathbf{D}_n^{(1)}$ is chosen so that

$$\left|\mathbb{E}D^2f(\mathbf{Y}_n)\left[\left(0,\mathbf{D}_n^{(2,1)}\right),\left(0,\mathbf{D}_n^{(2,1)}\right)\right]\right|$$

is small and this choice is made in an arbitrary way.

6.5. Distance from the continuous process

We now establish a bound on the speed of convergence of $\mathbf{Y}_n$ to a continuous Gaussian process whose covariance is the limit of the covariance of $\mathbf{D}_n$. We do this by bounding the distance between $\mathbf{D}_n$ and the continuous process via the Brownian modulus of continuity and using Theorem 6.2.

Theorem 6.4. *Let $\mathbf{Y}_n$ be defined as in Subsection 6.1 and let $\mathbf{Z} = (\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)})$ be defined by:*

$$\begin{cases} \mathbf{Z}^{(1)}(t) = \frac{\sqrt{p(1-p)}}{\sqrt{2+8p^2}} t\mathbf{B}_1(t^2) + \frac{p\sqrt{2p(1-p)}}{\sqrt{1+4p^2}} t\mathbf{B}_2(t^2), \\ \mathbf{Z}^{(2)}(t) = \frac{p\sqrt{2p(1-p)}}{\sqrt{1+4p^2}} t\mathbf{B}_1(t^2) + \frac{2p^2\sqrt{2p(1-p)}}{\sqrt{1+4p^2}} t\mathbf{B}_2(t^2) \end{cases},$$

where $\mathbf{B}_1, \mathbf{B}_2$ are independent standard Brownian Motions. Then, for any $g \in M^2$:

$$|\mathbb{E}g(\mathbf{Y}_n) - \mathbb{E}g(\mathbf{Z})| \leq \|g\|_{M^2} \left(913 n^{-1/2}\sqrt{\log n} + 112 n^{-1/2}\right).$$

Remark 6.5. *Theorem 6.4, together with Proposition 2.3, implies that $\mathbf{Y}_n$ converges to $\mathbf{Z}$ in distribution with respect to the Skorokhod and uniform topologies.*

In **Step 1** of the proof of Theorem 6.4, we provide a coupling between $\mathbf{D}_n$ and i.i.d standard Brownian Motions. Using those Brownian Motions, we construct a process $\mathbf{Z}_n$ having the same distribution as $\mathbf{D}_n$. In **Step 2** we couple $\mathbf{Z}_n$ and $\mathbf{Z}$ and bound the first two moments of the supremum distance between them, using the Brownian modulus of continuity. In **Step 3** we use those bounds together with the Mean Value Theorem to obtain Theorem 6.4.

Proof of Theorem 6.4.
Step 1. Let $\mathbf{B}_1$, $\mathbf{B}_2$, $\mathbf{B}_3$, $\mathbf{B}_4$, $\mathbf{B}_5$ be i.i.d. standard Brownian Motions and let $\mathbf{Z}_n = \left(\mathbf{Z}_n^{(1)}, \mathbf{Z}_n^{(2)}\right)$ be defined by:

$$\begin{aligned} \text{A)} \quad \mathbf{Z}_n^{(1)}(t) =& \frac{(\lfloor nt\rfloor - 2)\sqrt{p(1-p)}}{n^2\sqrt{2+8p^2}} \mathbf{B}_1\left(\lfloor nt\rfloor(\lfloor nt\rfloor - 1)\right) \\ &+ \frac{(\lfloor nt\rfloor - 2)p\sqrt{2p(1-p)}}{n^2\sqrt{1+4p^2}} \mathbf{B}_2\left(\lfloor nt\rfloor(\lfloor nt\rfloor - 1)\right); \\ \text{B)} \quad \mathbf{Z}_n^{(2)}(t) =& \frac{(\lfloor nt\rfloor - 2)p\sqrt{2p(1-p)}}{n^2\sqrt{1+4p^2}} \mathbf{B}_1\left(\lfloor nt\rfloor(\lfloor nt\rfloor - 1)\right) \\ &+ \frac{(\lfloor nt\rfloor - 2)2p^2\sqrt{2p(1-p)}}{n^2\sqrt{1+4p^2}} \mathbf{B}_2\left(\lfloor nt\rfloor(\lfloor nt\rfloor - 1)\right) \\ &+ \frac{\lfloor nt\rfloor - 2}{n^{5/2}} \mathbf{B}_3\left(\lfloor nt\rfloor(\lfloor nt\rfloor - 1)\right) + \frac{p(1-p)}{\sqrt{2}n^2} \mathbf{B}_4\left(\lfloor nt\rfloor^2(\lfloor nt\rfloor - 1)\right) \\ &+ \frac{\sqrt{2p^3(1-p)}}{n^2} \mathbf{B}_5(1). \end{aligned}$$

Now, note that $\left(\mathbf{D}_n^{(1)}, \mathbf{D}_n^{(2)}\right) \stackrel{\mathcal{D}}{=} \left(\mathbf{Z}_n^{(1)}, \mathbf{Z}_n^{(2)}\right)$. To see this, observe that for all

$u, t \in [0,1]$,

$$\text{A)}\quad \mathbb{E}\mathbf{D}_n^{(1)}(t)\mathbf{D}_n^{(1)}(u) = (\lfloor nt\rfloor - 2)(\lfloor nu\rfloor - 2)\lfloor n(t\wedge u)\rfloor(\lfloor n(t\wedge u)\rfloor - 1)\frac{p(1-p)}{2n^4}$$
$$= \mathbb{E}\mathbf{Z}_n^{(1)}(t)\mathbf{Z}_n^{(1)}(u);$$
$$\text{B)}\quad \mathbb{E}\mathbf{D}_n^{(2)}(t)\mathbf{D}_n^{(2)}(u) = \mathbb{E}\mathbf{Z}_n^{(2)}(t)\mathbf{Z}_n^{(2)}(u);$$
$$\text{C)}\quad \mathbb{E}\mathbf{D}_n^{(1)}(t)\mathbf{D}_n^{(2)}(u) = (\lfloor nt\rfloor - 2)(\lfloor nu\rfloor - 2)\lfloor n(t\wedge u)\rfloor(\lfloor n(t\wedge u)\rfloor - 1)\frac{p^2(1-p)}{n^4}$$
$$= \mathbb{E}\mathbf{Z}_n^{(1)}(t)\mathbf{Z}_n^{(2)}(u), \tag{6.7}$$

where B) is proved in Lemma 7.4, in the appendix.

Step 2. We let $\mathbf{Z}$ and $\mathbf{Z}_n$ be coupled in such a way that $\mathbf{Z}$ is constructed as in Theorem 6.4, using the same Brownian Motions $\mathbf{B}_1, \mathbf{B}_2$, as the ones used in the construction of $\mathbf{Z}_n$. In Lemma 7.5, proved in the appendix, we derive bounds for moments of the supremum distance between $\mathbf{Z}$ and $\mathbf{Z}_n$:

$$\mathbb{E}\left\|\mathbf{Z}_n - \mathbf{Z}\right\| \le \frac{12}{n^{1/2}} + \frac{51\sqrt{\log n}}{\sqrt{n}};$$
$$\mathbb{E}\left\|\mathbf{Z}_n - \mathbf{Z}\right\|^2 \le \frac{121}{n} + \frac{743\log n}{n};$$
$$\mathbb{E}\|\mathbf{Z}\|^2 \le 5. \tag{6.8}$$

Step 3. We note that $\|Dg(w)\| \le \|g\|_{M^2}(1+\|w\|)$ and therefore, by (6.8):

$$\begin{aligned}
|\mathbb{E}g(\mathbf{Z}) - \mathbb{E}g(\mathbf{D}_n)| &\overset{\text{MVT}}{\le} \mathbb{E}\left[\sup_{c\in[0,1]}\|Dg(\mathbf{Z}+c(\mathbf{Z}_n-\mathbf{Z}))\|\,\|\mathbf{Z}-\mathbf{Z}_n\|\right] \\
&\le \|g\|_{M^2}\mathbb{E}\left[\sup_{c\in[0,1]}(1+\|\mathbf{Z}+c(\mathbf{Z}_n-\mathbf{Z})\|)\,\|\mathbf{Z}-\mathbf{Z}_n\|\right] \\
&\le \|g\|_{M^2}\mathbb{E}\left[\|\mathbf{Z}-\mathbf{Z}_n\| + \|\mathbf{Z}\|\|\mathbf{Z}-\mathbf{Z}_n\| + \|\mathbf{Z}-\mathbf{Z}_n\|^2\right] \\
&\le \|g\|_{M^2}\left[\mathbb{E}\|\mathbf{Z}-\mathbf{Z}_n\| + \sqrt{\mathbb{E}\|\mathbf{Z}\|^2\mathbb{E}\|\mathbf{Z}-\mathbf{Z}_n\|^2} + \mathbb{E}\|\mathbf{Z}-\mathbf{Z}_n\|^2\right] \\
&\le \|g\|_{M^2}\left(901n^{-1/2} + 112n^{-1/2}\sqrt{\log n}\right),
\end{aligned}$$

which, together with Theorem 6.2 gives the desired result. □

Remark 6.6. *The representation of* $\mathbf{Z}$ *in terms of two independent Brownian Motions comes from a careful analysis of the limiting covariance of* $\mathbf{D}_n$*. Indeed, (6.7) provides an explicit derivation of the covariance, which converges to the covariance of* $\mathbf{Z}$*.*

Acknowledgements

The author would like to thank Gesine Reinert for many helpful discussions and comments on the early versions of this work. The author is also grateful to Andrew Barbour, Christian Döbler, Ivan Nourdin, Giovanni Peccati and Yvik Swan for many interesting discussions.

References

[Bar90] A.D. Barbour. Stein's Method for Diffusion Approximation. *Probability Theory and Related Fields*, 84:297–322, 1990.

[BDM18] E. Besançon, L. Decreusefond, and P. Moyal. Stein's method for diffusive limit of Markov processes. arXiv:1805.01691, 2018.

[Bil68] P. Billingsley. *Convergence of Probability Measures*. Wiley, New York, 1968.

[BJ09] A.D. Barbour and S. Janson. A functional combinatorial central limit theorem. *Electronic Journal of Probability*, 14(81):2352–2370, 2009.

[Bol84] E. Bolthausen. An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete*, 66(3):379–386, 1984.

[CD13] L. Coutin and L. Decreusefond. Stein's method for Brownian Approximations. *Communications on Stochastic Analysis*, 7(3):349–372, 2013.

[CDM05] S. Chatterjee, P. Diaconis, and E. Meckes. Exchangeable pairs and poisson approximation. *Probab. Surveys*, 2:64–106, 2005.

[CF15] Louis H.Y. Chen and Xiao Fang. On the error bound in a combinatorial central limit theorem. *Bernoulli*, 21(1):335–359, 02 2015.

[CFR11] S. Chatterjee, J. Fulman, and A. Röllin. Exponential approximation by stein's method and spectral graph theory. *ALEA Lat. Am. J. Probab. Math. Stat*, 2011.

[CM08] S. Chatterjee and E. Meckes. Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 4:257–283, 2008.

[Döb15] Ch. Döbler. Stein's method of exchangeable pairs for the beta distribution and generalizations. *Electron. J. Probab.*, 20:34 pp., 2015.

[Dud02] R. M. Dudley. *Real Analysis and Probability:*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2002.

[EK86] S.N. Ethier and T.G. Kurtz. *Markov processes: characterization and convergence*. Wiley, New York, 1986.

[FN10] M. Fischer and G. Nappo. On the Moments of the Modulus of Continuity of Ito Processes. *Stochastic Analysis and Applications*, 28(1):103–122, 2010.

[Gol05] L. Goldstein. Berry-Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *J. Appl. Probab.*, 42(3):661–683, 2005.

[HC78] S.-T. Ho and L.H. Y. Chen. An l_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.*, 6(2):231–249, 04 1978.

[JN91] S. Janson and K. Nowicki. The asymptotic distributions of generalized U-statistics with applications to random graphs. *Probability Theory and Related Fields*, 90(3):341–375, Sep 1991.

[Kas17a] M.J. Kasprzak. Diffusion approximations via Stein's method and time changes. arXiv:1701.07633, 2017.

[Kas17b] M.J. Kasprzak. Stein's method for multivariate Brownian approximations of sums under dependence. arXiv:1708.02521, 2017.

[KDV17] M.J. Kasprzak, A. B. Duncan, and S.J. Vollmer. Note on A. Barbour's paper on Stein's method for diffusion approximations. *Electron. Commun. Probab.*, 22(23):1–8, 2017.

[Mec09] E. Meckes. *On Stein's method for multivariate normal approximation*, volume Volume 5 of *Collections*, pages 153–178. Institute of Mathematical Statistics, Beachwood, Ohio, USA, 2009.

[NP12] I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus*. Cambridge tracts in Mathematics. Cambridge University Press, 2012.

[NR12] K. Neammanee and N. Rerkruthairat. An improvement of a uniform bound on a combinatorial central limit theorem. *Comm. Statist. Theory Methods*, 41(9):1590–1602, 2012.

[Röl07] A. Röllin. Translated poisson approximation using exchangeable pair couplings. *The Annals of Applied Probability*, 17(5/6):1596–1614, 2007.

[RR97] Y. Rinott and V. Rotar. On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.*, 7(4):1080–1105, 11 1997.

[RR09] G. Reinert and A. Röllin. Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *The Annals of Probability*, 37(6):2150–2173, 2009.

[RR10] G. Reinert and A. Röllin. Random subgraph counts and u-statistics: Multivariate normal approximation via exchangeable pairs and embedding. *Journal of Applied Probability*, 47(2):378–393, 2010.

[Ste72] Ch. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, 2:583–602, 1972.

[Ste86] Ch. Stein. *Approximate Computation of Expectations*. Institute of Mathematical Statistics Lecture Notes, Monograph Series, 7. Hayward, Calif. Institute of Mathematical Statistics, 1986.

[WW44] A. Wald and J. Wolfowitz. Statistical tests based on permutations of the observations. *Ann. Math. Statistics*, 15:358–372, 1944.

7. Appendix - technical details of the proofs of Theorems 5.1, 6.2 and 6.4

7.1. Technical details of the proof of Theorem 5.1

Lemma 7.1. *In the setup of Theorem 6.2 and for ϵ_2 defined by Theorem 4.1,*

$$\begin{aligned}\epsilon_2 =& \left|\mathbb{E}D^2 f(\mathbf{Y}_n)\left[(\mathbf{Y}_n-\mathbf{Y}_n')\Lambda_n, \mathbf{Y}_n-\mathbf{Y}_n'\right] - \mathbb{E}D^2 f(\mathbf{Y}_n)[\mathbf{D}_n,\mathbf{D}_n]\right|\\ \leq& A+B,\end{aligned}$$

for

$$A = \Bigg| \frac{1}{n(n-1)s_n^2} \sum_{\substack{1\le i,j,k,l\le n\\ i\neq j, k\neq l}} \mathbb{E}\left\{ \left[\frac{(X_{ik}-X_{il})^2}{2n} - \frac{\hat{Z}_i^2}{n-1} \right] \right.$$

$$\left. \cdot \left(D^2 f(Y_{n,ijkl}) - D^2 f\left(Y_n^{ijkl}\right)\right) \left[\mathbb{1}_{[i/n,1]} \mathbb{1}_{[i/n,1]} \right] \right\}$$

$$+ \frac{1}{n(n-1)s_n^2} \sum_{\substack{1\le i,j,k,l\le n\\ i\neq j, k\neq l}} \mathbb{E}\left\{ \left[\frac{(X_{ik}-X_{il})(X_{jl}-X_{jk})}{2n} - \hat{Z}_i\hat{Z}_j \right] \right.$$

$$\left. \cdot \left(D^2 f(Y_{n,ijkl}) - D^2 f\left(Y_n^{ijkl}\right)\right) \left[\mathbb{1}_{[i/n,1]}, \mathbb{1}_{[j/n,1]} \right] \right\} \Bigg|,$$

$$B = \Bigg| \frac{1}{n(n-1)s_n^2} \sum_{\substack{1\le i,j,k,l\le n\\ i\neq j, k\neq l}} \mathbb{E}\left\{ \left[\frac{(X_{ik}-X_{il})^2}{2n} - \frac{\hat{Z}_i^2}{n-1} \right] D^2 f\left(Y_n^{ijkl}\right) \left[\mathbb{1}_{[i/n,1]} \mathbb{1}_{[i/n,1]} \right] \right\}$$

$$+ \frac{1}{n(n-1)s_n^2} \sum_{\substack{1\le i,j,k,l\le n\\ i\neq j, k\neq l}} \mathbb{E}\left\{ \left[\frac{(X_{ik}-X_{il})(X_{jl}-X_{jk})}{2n} - \hat{Z}_i\hat{Z}_j \right] \right.$$

$$\left. \cdot D^2 f\left(Y_n^{ijkl}\right) \left[\mathbb{1}_{[i/n,1]}, \mathbb{1}_{[j/n,1]} \right] \right\} \Bigg|.$$

Proof. Note that

$$\begin{aligned} \epsilon_2 &= \left| \mathbb{E} D^2 f(\mathbf{Y}_n) \left[(\mathbf{Y}_n - \mathbf{Y}_n')\Lambda_n, \mathbf{Y}_n - \mathbf{Y}_n' \right] - \mathbb{E} D^2 f(\mathbf{Y}_n)[\mathbf{D}_n, \mathbf{D}_n] \right| \\ &= \left| \frac{n-1}{4} \mathbb{E} D^2 f(\mathbf{Y}_n)[\mathbf{Y}_n - \mathbf{Y}_n', \mathbf{Y}_n - \mathbf{Y}_n'] - \mathbb{E} D^2 f(\mathbf{Y}_n)[\mathbf{D}_n, \mathbf{D}_n] \right| \end{aligned} \quad (7.1)$$

and

$$\begin{aligned} &\frac{n-1}{4} \mathbb{E} D^2 f(\mathbf{Y}_n)[\mathbf{Y}_n - \mathbf{Y}_n', \mathbf{Y}_n - \mathbf{Y}_n'] - \mathbb{E} D^2 f(\mathbf{Y}_n)[\mathbf{D}_n, \mathbf{D}_n] \\ =&\frac{1}{2ns_n^2} \sum_{i,j=1}^n \mathbb{E}\left\{ (X_{i\pi(i)} - X_{i\pi(j)})^2 D^2 f(\mathbf{Y}_n) \left[\mathbb{1}_{[i/n,1]} \mathbb{1}_{[i/n,1]} \right] \right\} \\ &+ \frac{1}{2ns_n^2} \sum_{i,j=1}^n \mathbb{E}\left\{ (X_{i\pi(i)} - X_{i\pi(j)})(X_{j\pi(j)} - X_{j\pi(i)}) D^2 f(\mathbf{Y}_n) \left[\mathbb{1}_{[i/n,1]}, \mathbb{1}_{[j/n,1]} \right] \right\} \\ &- \mathbb{E} D^2 f(\mathbf{Y}_n)[\mathbf{D}_n, \mathbf{D}_n] \end{aligned}$$

$$=\frac{1}{2n^2(n-1)s_n^2}\sum_{\substack{1\le i,j,k,l\le n\\ i\neq j,k\neq l}}\mathbb{E}\left\{(X_{ik}-X_{il})^2\cdot D^2f(\mathbf{Y}_n)\left[\mathbb{1}_{[i/n,1]}\mathbb{1}_{[i/n,1]}\right]\middle|\,\pi(i)=k,\pi(j)=l\right\}$$

$$+\frac{1}{n(n-1)s_n^2}\sum_{\substack{1\le i,j,k,l\le n\\ i\neq j,k\neq l}}\mathbb{E}\left\{\left[\frac{(X_{ik}-X_{il})(X_{jl}-X_{jk})}{2n}\right]\right.$$

$$\left.\cdot\, D^2f(\mathbf{Y}_n)\left[\mathbb{1}_{[i/n,1]},\mathbb{1}_{[j/n,1]}\right]\,\middle|\,\pi(i)=k,\pi(j)=l\right\}$$

$$-\frac{1}{s_n^2}\sum_{1\le i\neq j\le n}\mathbb{E}[\hat{Z}_i\hat{Z}_j]\mathbb{E}D^2f(\mathbf{Y}_n)[\mathbb{1}_{[i/n,1]},\mathbb{1}_{[j/n,1]}]$$

$$-\frac{1}{(n-1)s_n^2}\sum_{1\le i\neq j\le n}\mathbb{E}[\hat{Z}_i^2]\mathbb{E}D^2f(\mathbf{Y}_n)[\mathbb{1}_{[i/n,1]},\mathbb{1}_{[i/n,1]}]$$

$$=\frac{1}{2n^2(n-1)s_n^2}\sum_{\substack{1\le i,j,k,l\le n\\ i\neq j,k\neq l}}\mathbb{E}\left\{(X_{ik}-X_{il})^2D^2f(Y_{n,ijkl})\left[\mathbb{1}_{[i/n,1]}\mathbb{1}_{[i/n,1]}\right]\right\}$$

$$+\frac{1}{n(n-1)s_n^2}\sum_{\substack{1\le i,j,k,l\le n\\ i\neq j,k\neq l}}\mathbb{E}\left\{\frac{(X_{ik}-X_{il})(X_{jl}-X_{jk})}{2n}D^2f(Y_{n,ijkl})\left[\mathbb{1}_{[i/n,1]},\mathbb{1}_{[j/n,1]}\right]\right\}$$

$$-\frac{1}{n(n-1)s_n^2}\sum_{\substack{1\le i,j,k,l\le n\\ i\neq j,k\neq l}}\mathbb{E}[\hat{Z}_i\hat{Z}_j]\mathbb{E}D^2f(Y_{n,ijkl})[\mathbb{1}_{[i/n,1]},\mathbb{1}_{[j/n,1]}]$$

$$-\frac{1}{n(n-1)^2s_n^2}\sum_{\substack{1\le i,j,k,l\le n\\ i\neq j,k\neq l}}\mathbb{E}[\hat{Z}_i^2]\mathbb{E}D^2f(Y_{n,ijkl})[\mathbb{1}_{[i/n,1]},\mathbb{1}_{[i/n,1]}]. \tag{7.2}$$

Now, the lemma follows by taking the absolute value in (7.2) and combining it with (7.1). ☐

Lemma 7.2. *For A of (5.9),*

$$A\le\frac{\|g\|_{M^1}}{n^3(n-1)s_n^3}\sum_{1\le i,j,k,l,u\le n}\left\{\mathbb{E}|X_{ik}|^3+5\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|^2+7\mathbb{E}|X_{ik}|^2\mathbb{E}|X_{jl}|\right.$$
$$+5\mathbb{E}|X_{ik}|^2\mathbb{E}|X_{jk}|+16\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|\mathbb{E}|X_{jl}|+2\mathbb{E}|X_{iu}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|$$
$$+4\mathbb{E}|X_{iu}|\mathbb{E}|X_{il}|\mathbb{E}|X_{jk}|+6\mathbb{E}|X_{uk}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{jl}|+2\mathbb{E}|X_{uk}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{jk}|$$
$$\left.+\frac{1}{n}\left(2\mathbb{E}\,|X_{ik}|+2\mathbb{E}\,|X_{j,l}|+2\mathbb{E}|X_{uk}|+2\mathbb{E}|X_{ul}|\right)\cdot\sum_{r=1}^n\left(\mathbb{E}|X_{ir}|^2+|c_{ir}c_{jr}|\right)\right\}.$$

Proof. Let us adopt the notation of the proof of Theorem 5.1. Define index sets $\mathcal{I}=\{i,j,\pi^{-1}(k),\pi^{-1}(l)\}$ and $\mathcal{J}=\{k,l,\pi(i),\pi(j)\}$. Then, letting $\mathbf{S}=$

$\frac{1}{s_n}\sum_{i'\notin\mathcal{I}} X_{i'\pi(i')}\mathbb{1}_{[i'/n,1]}$, we can write:

$$\mathbf{Y}_{n,ijkl}=\mathbf{S}+\frac{1}{s_n}\sum_{i'\in\mathcal{I}}X_{i'\pi_{ijkl}(i')}\mathbb{1}_{[i'/n,1]},\qquad \mathbf{Y}_n^{ijkl}=\mathbf{S}+\frac{1}{s_n}\sum_{i'\in\mathcal{I}}X^{ijkl}_{i'\pi(i')}\mathbb{1}_{[i'/n,1]}.$$

Since $\mathbf{S}$ depends only on the components of $\mathbb{X}$ outside the square $\mathcal{I}\times\mathcal{J}$ and $\{\pi(i): i\notin\mathcal{I}\}$, $\mathbf{S}$ is independent of:

$$\left\{X_{il},X_{jk},X_{ik},X_{jl},\sum_{i'\in\mathcal{I}}X_{i'\pi_{ijkl}(i')},\sum_{i'\in\mathcal{I}}X^{ijkl}_{i'\pi(i')}\right\},$$

given $\pi^{-1}(k),\pi^{-1}(l),\pi(i),\pi(j)$.

Note that, by Proposition 3.2,

$$\begin{aligned}
A\leq&\frac{\|g\|_{M^1}}{n(n-1)s_n^2}\sum_{\substack{1\leq i,j,k,l\leq n\\ i\neq j,k\neq l}}\mathbb{E}\left\{\left\|Y_{n,ijkl}-\mathbf{Y}_n^{ijkl}\right\|\left(\left|\frac{(X_{ik}-X_{il})^2}{2n}-\frac{\mathbb{E}\hat{Z}_i^2}{n-1}\right|\right.\right.\\
&\left.\left.+\left|\frac{(X_{ik}-X_{il})(X_{jl}-X_{jk})}{2n}-\mathbb{E}(\hat{Z}_i\hat{Z}_j)\right|\right)\right\}\\
\leq&\frac{\|g\|_{M^1}}{n(n-1)s_n^3}\sum_{\substack{1\leq i,j,k,l\leq n\\ i\neq j,k\neq l}}\sum_{i'\in\mathcal{I}}\mathbb{E}\left\{\left|X_{i',\pi_{ijkl}(i')}-X^{ijkl}_{i'\pi(i')}\right|\right.\\
&\left.\cdot\left(\left|\frac{(X_{ik}-X_{il})^2}{2n}-\frac{\mathbb{E}\hat{Z}_i^2}{n-1}\right|+\left|\frac{(X_{ik}-X_{il})(X_{jl}-X_{jk})}{2n}-\mathbb{E}(\hat{Z}_i\hat{Z}_j)\right|\right)\right\}\\
\leq&\frac{\|g\|_{M^1}}{n(n-1)s_n^3}\sum_{\substack{1\leq i,j,k,l\leq n\\ i\neq j,k\neq l}}\mathbb{E}\left\{\left(\left|X_{ik}-X^{ijkl}_{i,\pi(i)}\right|+\left|X_{j,l}-X^{ijkl}_{j,\pi(j)}\right|+\left|X_{i,k}-X^{ijkl}_{\pi^{-1}(k),k}\right|\right.\right.\\
&\left.\left.+\left|X_{j,l}-X^{ijkl}_{\pi^{-1}(l),l}\right|\right)\left(\left|\frac{(X_{ik}-X_{il})^2}{2n}-\frac{\mathbb{E}\hat{Z}_i^2}{n-1}\right|+\left|\frac{(X_{ik}-X_{il})(X_{jl}-X_{jk})}{2n}-\mathbb{E}(\hat{Z}_i\hat{Z}_j)\right|\right)\right\}\\
\leq&\frac{\|g\|_{M^1}}{2n(n-1)^2s_n^3}\sum_{\substack{1\leq i,j,k,l\leq n\\ i\neq j,k\neq l}}\mathbb{E}\left\{\left(2\left|X_{ik}\right|+\left|X^{ijkl}_{i,\pi(i)}\right|+2\left|X_{j,l}\right|+\left|X^{ijkl}_{j,\pi(j)}\right|+\left|X^{ijkl}_{\pi^{-1}(k),k}\right|\right.\right.\\
&\left.+\left|X^{ijkl}_{\pi^{-1}(l),l}\right|\right)\left(\left|X_{ik}\right|^2+\left|X_{il}\right|^2+2\left|X_{ik}X_{il}\right|+2\left|\hat{Z}_i\right|^2+\left|X_{ik}X_{jl}\right|+\left|X_{ik}X_{jk}\right|\right.\\
&\left.\left.+\left|X_{il}X_{jl}\right|+\left|X_{il}X_{jk}\right|+2(n-1)\left|\mathbb{E}(\hat{Z}_i\hat{Z}_j)\right|\right)\right\}\\
\leq&\frac{\|g\|_{M^1}}{2n^2(n-1)s_n^3}\sum_{\substack{1\leq i,j,k,l\leq n\\ i\neq j,k\neq l}}\mathbb{E}\left\{\left(\left|X_{ik}\right|+\left|X_{j,l}\right|\right)\left(\left|X_{ik}\right|^2+\left|X_{il}\right|^2+2\left|X_{ik}X_{il}\right|+\frac{2}{n}\sum_{r=1}^n\mathbb{E}|X_{ir}|^2\right.\right.\\
&\left.\left.+\left|X_{ik}X_{jl}\right|+\left|X_{ik}X_{jk}\right|+\left|X_{il}X_{jl}\right|+\left|X_{il}X_{jk}\right|+\frac{2}{n}\sum_{r=1}^n\left|c_{ir}c_{jr}\right|\right)\right\}
\end{aligned}$$

$$+\frac{\|g\|_{M^1}}{n^3(n-1)s_n^3}\sum_{\substack{1\leq i,j,k,l,u\leq n\\ i\neq j,k\neq l}}\left(\mathbb{E}|X_{i,u}|+\mathbb{E}|X_{j,u}|+\mathbb{E}|X_{uk}|+\mathbb{E}|X_{u,l}|\right)\mathbb{E}\left\{|X_{ik}|^2+|X_{il}|^2\right.$$

$$\left.+2\,|X_{ik}X_{il}|+\frac{1}{n}\sum_{r=1}^n\mathbb{E}|X_{ir}|^2+|X_{ik}X_{jl}|+|X_{ik}X_{jk}|+|X_{il}X_{jl}|+|X_{il}X_{jk}|+\frac{1}{n}\sum_{r=1}^n|c_{ir}c_{jr}|\right\}$$

$$\leq\frac{\|g\|_{M^1}}{n^3(n-1)s_n^3}\sum_{1\leq i,j,k,l,u\leq n}\left\{\mathbb{E}|X_{ik}|^3+5\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|^2+7\mathbb{E}|X_{ik}|^2\mathbb{E}|X_{jl}|\right.$$

$$+5\mathbb{E}|X_{ik}|^2\mathbb{E}|X_{jk}|+16\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|\mathbb{E}|X_{jl}|+2\mathbb{E}|X_{iu}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{il}|$$

$$+4\mathbb{E}|X_{iu}|\mathbb{E}|X_{il}|\mathbb{E}|X_{jk}|+6\mathbb{E}|X_{uk}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{jl}|+2\mathbb{E}|X_{uk}|\mathbb{E}|X_{ik}|\mathbb{E}|X_{jk}|$$

$$\left.+\frac{1}{n}\left(2\mathbb{E}\,|X_{ik}|+2\mathbb{E}\,|X_{j,l}|+2\mathbb{E}|X_{uk}|+2\mathbb{E}|X_{ul}|\right)\cdot\sum_{r=1}^n\left(\mathbb{E}|X_{ir}|^2+|c_{ir}c_{jr}|\right)\right\}.$$

which finishes the proof. □

7.2. Technical details of the proof of Theorem 6.2

Lemma 7.3. *For $S_i, i=1,\cdots,7$ of (6.4), we have the following estimates:*

$$S_1\leq\frac{\sqrt{5}\|g\|_{M^2}}{12n},\quad S_2\leq\frac{\sqrt{5}\|g\|_{M^2}}{12n},\quad S_3\leq\frac{\sqrt{178}\|g\|_{M^2}}{6n},\quad S_4\leq\frac{\sqrt{178}\|g\|_{M^2}}{6n}$$
$$S_5\leq\frac{\sqrt{178}\|g\|_{M^2}}{12n},\quad S_6\leq\frac{\sqrt{612}\|g\|_{M^2}}{6n},\quad S_7\leq\frac{\|g\|_{M^2}}{n}.$$

Proof. For S_1, for fixed $i,j\in\{1,\cdots,n\}$, let $\mathbf{Y}_n^{ij}$ be equal to $\mathbf{Y}_n$ except for the fact that I_{ij} is replaced by an independent copy, i.e. for all $t\in[0,1]$ let:

$$\mathbf{T}_n^{ij}(t)=\mathbf{T}_n(t)-\frac{\lfloor nt\rfloor-2}{n^2}\left(I_{ij}-I'_{ij}\right)\mathbb{1}_{[i/n,1]\cap[j/n,1]}(t)$$
$$\mathbf{V}_n^{ij}(t)=\mathbf{V}_n(t)-\frac{1}{n^2}\sum_{k:k\neq i,j}\left(I_{ij}-I'_{ij}\right)\left(I_{jk}+I_{ik}\right)\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(t)$$

and let $\mathbf{Y}_n^{ij}(t)=\left(\mathbf{T}_n^{ij}(t)-\mathbb{E}\mathbf{T}_n(t),\mathbf{V}_n^{ij}(t)-\mathbb{E}\mathbf{V}_n(t)\right)$.

By noting that the mean zero $Z_i^{(1)}$ and $Z_j^{(1)}$ are independent for $i\neq j$, we obtain:

$$S_1=\left|\frac{n(n-1)}{8}\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbf{T}'_n)(2,0),(\mathbf{T}_n-\mathbf{T}'_n)(1,0)\right]\right.$$
$$-\sum_{j,k=1}^n\mathbb{E}D^2f(\mathbf{Y}_n)\left[\sum_{i=1}^n Z_{i,k}^{(1)}(\lfloor n\cdot\rfloor-2)(1,0)\mathbb{1}_{[i/n,1]\cap[k/n,1]},\right.$$
$$\left.\left.\sum_{i=1}^n Z_{i,j}^{(1)}(\lfloor n\cdot\rfloor-2)(1,0)\mathbb{1}_{[i/n,1]\cap[j/n,1]}\right]\right|$$

$$=\left|\frac{1}{4n^4}\sum_{1\leq i\neq j\leq n}\mathbb{E}\left\{(I_{i,j}-2pI_{i,j}+p)\right.\right.$$
$$\left.\cdot D^2f(\mathbf{Y}_n)\left[(\lfloor n\cdot\rfloor-2)\mathbb{1}_{[i/n,1]\cap[j/n,1]}(1,0),(\lfloor n\cdot\rfloor-2)\mathbb{1}_{[i/n,1]\cap[j/n,1]}(1,0)\right]\right\}$$
$$-\sum_{i,j=1}^{n}\left\{\mathbb{E}\left(Z_{i,j}^{(1)}\right)^2\right.$$
$$\left.\left.\cdot\,\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\lfloor n\cdot\rfloor-2)(1,0)\mathbb{1}_{[i/n,1]\cap[j/n,1]},(\lfloor n\cdot\rfloor-2)(1,0)\mathbb{1}_{[i/n,1]\cap[j/n,1]}\right]\right\}\right|$$
$$=\left|\sum_{1\leq i\neq j\leq n}\mathbb{E}\left\{\left(\frac{1}{4n^4}(I_{i,j}-2pI_{i,j}+p)-\mathbb{E}\left(Z_{i,j}^{(1)}\right)^2\right)\right.\right.$$
$$\left.\left.\cdot\, D^2f(\mathbf{Y}_n)\left[(\lfloor n\cdot\rfloor-2)(1,0)\mathbb{1}_{[i/n,1]\cap[j/n,1]},(\lfloor n\cdot\rfloor-2)(1,0)\mathbb{1}_{[i/n,1]\cap[j/n,1]}\right]\right\}\right|$$
$$=\left|\sum_{1\leq i\neq j\leq n}\mathbb{E}\left\{\frac{1}{4n^4}(I_{i,j}-2pI_{i,j}+p)\right.\right.$$
$$\left.\left.\cdot\left(D^2f(\mathbf{Y}_n)-D^2f(\mathbf{Y}_n^{ij})\right)\left[(\lfloor n\cdot\rfloor-2)(1,0)\mathbb{1}_{[i/n,1]\cap[j/n,1]},(\lfloor n\cdot\rfloor-2)(1,0)\mathbb{1}_{[i/n,1]\cap[j/n,1]}\right]\right\}\right|$$
$$\leq\frac{\|g\|_{M^2}}{12n^2}\sum_{1\leq i\neq j\leq n}\mathbb{E}\left|(I_{i,j}-2pI_{i,j}+p)\right|\left\|\mathbf{Y}_n-\mathbf{Y}_n^{ij}\right\|,\tag{7.3}$$

where (7.3) follows from Proposition 3.2. Now,

$$\left\|\mathbf{Y}_n-\mathbf{Y}_n^{ij}\right\|\leq\frac{1}{n^2}\sqrt{(\lfloor n\cdot\rfloor-2)^2(I_{ij}-I'_{ij})^2+\left(\sum_{k:k\neq i,j}|I_{ij}-I'_{ij}|(I_{jk}+I_{ik})\right)^2}$$

and so, by (7.3),

$$S_1\leq\frac{\|g\|_{M^2}}{12n^4}\sum_{1\leq i\neq j\leq n}\mathbb{E}\left\{|I_{i,j}-2pI_{i,j}+p|\right.$$
$$\left.\cdot\sqrt{(n-2)^2(I_{ij}-I'_{ij})^2+\left(\sum_{k\neq i,j}|I_{ij}-I'_{ij}|(I_{jk}+I_{ik})\right)^2}\right\}$$
$$\leq\frac{\|g\|_{M^2}}{12n^3}\sum_{1\leq i\neq j\leq n}\mathbb{E}\left\{|I_{i,j}-2pI_{i,j}+p|\cdot\sqrt{(I_{ij}-I'_{ij})^2+\left(|I_{ij}-I'_{ij}|(I_{jk}+I_{ik})\right)^2}\right\}$$

$$\leq \frac{\sqrt{5}\|g\|_{M^2}}{12n}, \tag{7.4}$$

where the last inequality holds because $|I_{ij} - 2pI_{ij} + p| \leq 1$, $|I_{ij} - I'_{ij}| \leq 1$ and $I_{jk} + I_{ik} \leq 2$ for all $k \in \{1, \cdots, n\}$.

Similarly, for S_2:

$$\begin{aligned}
S_2 = &\Bigg| \frac{n(n-1)}{8} \mathbb{E}\left\{ D^2 f(\mathbf{Y}_n) \left[(\mathbf{T}_n - \mathbf{T}'_n)(0, 2p), (\mathbf{T}_n - \mathbf{T}'_n)(1,0) \right] \right\} \\
&- 2 \sum_{j,k=1}^{n} \mathbb{E} D^2 f(\mathbf{Y}_n) \left[\sum_{i=1}^{n} Z_{i,k}^{(2,1)} (\lfloor n \cdot \rfloor - 2) \mathbb{1}_{[i/n,1] \cap [k/n,1]}(0,1), \right. \\
&\left. \sum_{i=1}^{n} Z_{i,j}^{(1)} (\lfloor n \cdot \rfloor - 2) \mathbb{1}_{[i/n,1] \cap [j/n,1]}(1,0) \right] \Bigg| \\
= &\Bigg| \frac{p}{4n^4} \sum_{1 \leq i \neq j \leq n} \mathbb{E} \{ (I_{i,j} - 2pI_{i,j} + p) \\
&\cdot D^2 f(\mathbf{Y}_n) \left[(\lfloor n \cdot \rfloor - 2) \mathbb{1}_{[i/n,1] \cap [j/n,1]}(0,1), (\lfloor n \cdot \rfloor - 2) \mathbb{1}_{[i/n,1] \cap [j/n,1]}(1,0) \right] \Big\} \\
&- \sum_{1 \leq i \neq j \leq n} \left\{ \mathbb{E} \left(Z_{i,j}^{(1)} Z_{i,j}^{(2,1)} \right) \right. \\
&\left. \cdot \mathbb{E} D^2 f(\mathbf{Y}_n) \left[(\lfloor n \cdot \rfloor - 2)(0,1) \mathbb{1}_{[i/n,1] \cap [j/n,1]}, (\lfloor n \cdot \rfloor - 2)(1,0) \mathbb{1}_{[i/n,1] \cap [j/n,1]} \right] \right\} \Bigg| \\
\leq &\frac{\sqrt{5} p \|g\|_{M^2}}{12n} \\
\leq &\frac{\sqrt{5}\|g\|_{M^2}}{12n}.
\end{aligned} \tag{7.5}$$

For S_3, let $\mathbf{Y}_n^{ijk}$ equal to $\mathbf{Y}_n$ except that I_{ij}, I_{jk}, I_{ik} are replaced by I'_{ij}, I'_{jk}, I'_{ik}, i.e. for all $t \in [0,1]$ let

$$\begin{aligned}
\mathbf{T}_n^{ijk}(t) = &\mathbf{T}_n(t) - \frac{\lfloor nt \rfloor - 2}{n^2} \left[(I_{ij} - I'_{ij}) \mathbb{1}_{[i/n,1] \cap [j/n,1]}(t) \right. \\
&\left. + (I_{jk} - I'_{jk}) \mathbb{1}_{[j/n,1] \cap [k/n,1]}(t) + (I_{ik} - I'_{ik}) \mathbb{1}_{[i/n,1] \cap [k/n,1]}(t) \right] \\
\mathbf{V}_n^{ijk}(t) = &\mathbf{V}_n(t) - \frac{1}{n^2} \sum_{l: l \neq i,j,k} \left[\left(I_{ij} - I'_{ij} \right) (I_{jl} + I_{il}) \mathbb{1}_{[i/n,1] \cap [j/n,1] \cap [l/n,1]}(t) \right. \\
&+ \left(I_{jk} - I'_{jk} \right) (I_{jl} + I_{kl}) \mathbb{1}_{[k/n,1] \cap [j/n,1] \cap [l/n,1]}(t) \\
&\left. + (I_{ik} - I'_{ik}) (I_{jl} + I_{il}) \mathbb{1}_{[i/n,1] \cap [k/n,1] \cap [l/n,1]}(t) \right] \\
&- \frac{1}{n^2} \left[(I_{ij} I_{jk} - I'_{ij} I'_{jk}) + (I_{ij} I_{ik} - I'_{ij} I'_{ik}) + (I_{ik} I_{jk} - I'_{ik} I'_{jk}) \right] \mathbb{1}_{[i/n,1] \cap [j/n,1] \cap [k/n,1]}(t).
\end{aligned} \tag{7.6}$$

Let $\mathbf{Y}_n^{ijk}(t) = \left(\mathbf{T}_n^{ijk}(t) - \mathbb{E}\mathbf{T}_n(t), \mathbf{V}_n^{ijk}(t) - \mathbb{E}\mathbf{V}_n(t)\right)$ for all $t \in [0,1]$. Note that

$$S_3 = \Bigg| \frac{n(n-1)}{8}\mathbb{E}D^2 f(\mathbf{Y}_n)\left[(\mathbf{T}_n - \mathbf{T}_n')(2,0), (\mathbf{V}_n - \mathbf{V}_n')(0,1)\right]$$

$$-\frac{4}{3}\mathbb{E}D^2 f(\mathbf{Y}_n)\left[\sum_{i,j=1}^n Z_{i,j}^{(1)}(\lfloor n\cdot\rfloor - 2)(1,0)\mathbb{1}_{[i/n,1]\cap[j/n,1]}, \sum_{i,j,k=1}^n Z_{i,j,l}^{(2,2)}(0,1)\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[l/n,1]}\right]\Bigg|$$

$$= \Bigg| \frac{1}{4n^4}\sum_{\substack{1\le i,j,k\le n\\ i,j,k \text{ distinct}}} \mathbb{E}\,\{(I_{ij} - I_{ij}')^2(I_{jk} + I_{ik})$$

$$\cdot D^2 f(\mathbf{Y}_n)\left[(\lfloor n\cdot\rfloor - 2)\mathbb{1}_{[i/n,1]\cap[j/n,1]}(1,0), \mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(0,1)\right]\}$$

$$-\frac{4}{3}\sum_{\substack{1\le i,j,k\le n\\ i,j,k \text{ distinct}}} \mathbb{E}\left\{Z_{i,j}^{(1)}Z_{i,j,k}^{(2,2)}D^2 f(\mathbf{Y}_n)\left[(\lfloor n\cdot\rfloor - 2)\mathbb{1}_{[i/n,1]\cap[j/n,1]}(1,0), \mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(0,1)\right]\right\}\Bigg|$$

$$= \Bigg| \sum_{\substack{1\le i,j,k\le n\\ i,j,k \text{ distinct}}} \mathbb{E}\left\{\left(\frac{1}{4n^4}(I_{ij} - 2pI_{ij} + p)(I_{jk} + I_{ik}) - \frac{4}{3}\mathbb{E}Z_{i,j}^{(1)}Z_{i,j,k}^{(2,2)}\right)\right.$$

$$\left.\cdot D^2 f(\mathbf{Y}_n)\left[(\lfloor n\cdot\rfloor - 2)\mathbb{1}_{[i/n,1]\cap[j/n,1]}(1,0), \mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(0,1)\right]\right\}\Bigg|$$

$$= \Bigg| \sum_{\substack{1\le i,j,k\le n\\ i,j,k \text{ distinct}}} \mathbb{E}\left\{\left(\frac{1}{4n^4}(I_{ij} - 2pI_{ij} + p)(I_{jk} + I_{ik})\right)\right.$$

$$\left.\cdot \left(D^2 f(\mathbf{Y}_n) - D^2 f(\mathbf{Y}_n^{ijk})\right)\left[(\lfloor n\cdot\rfloor - 2)\mathbb{1}_{[i/n,1]\cap[j/n,1]}(1,0), \mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(0,1)\right]\right\}\Bigg|$$

$$\le \frac{\|g\|_{M^2}}{12n^3}\sum_{\substack{1\le i,j,k\le n\\ i,j,k \text{ distinct}}} \mathbb{E}\left(I_{ij} - 2pI_{ij} + p\right)\left(I_{jk} + I_{ik}\right)\left\|\mathbf{Y}_n - \mathbf{Y}_n^{ijk}\right\|. \tag{7.7}$$

Now, by (7.6), we note that:

$$\|\mathbf{Y}_n - \mathbf{Y}_n^{ijk}\| \le \frac{1}{n^2}\left\{(n-2)^2(|I_{ij} - I_{ij}'| + |I_{jk} - I_{jk}'| + |I_{ik} - I_{ik}'|)^2\right.$$

$$+\left[\sum_{l:l\neq i,j,k}\left(|I_{ij} - I_{ij}'|(I_{jl} + I_{il}) + |I_{jk} - I_{jk}'|(I_{jl} + I_{kl}\right)\right.$$

$$
\begin{aligned}
&+|I_{ik}-I'_{ik}|(I_{jl}+I_{il})+|I_{ik}-I'_{ik}|(I_{jl}+I_{il}))\\
&+|I_{ij}I_{jk}-I'_{ij}I'_{jk}|+|I_{ij}I_{ik}-I'_{ij}I'_{ik}|+|I_{ij}I_{jk}-I'_{ij}I'_{jk}|\big]^2\Big\}^{1/2}\\
\leq&\frac{1}{n^2}\sqrt{9(n-2)^2+(8(n-3)+3)^2}\\
=&\frac{\sqrt{73n^2-372n+477}}{n^2},
\end{aligned}
$$

where the second inequality follows from the fact that for all $a,b,c\in\{1,\cdots,n\}$, $|I_{ab}-I'_{ab}|\leq 1$, $(I_{ab}+I_{bc})\leq 2$ and $|I_{ab}I_{bc}-I'_{ab}I'_{bc}|\leq 1$. Also, $(I_{jk}+I_{ik})\leq 2$ and $I_{ij}-2pI_{ij}+p\leq 1$. Therefore, by (7.7):

$$
\begin{aligned}
S_3\leq&\frac{\|g\|_M n(n-1)(n-2)\sqrt{73n^2-372n+477}}{6n^5}\\
\leq&\frac{\sqrt{178}\|g\|_{M^2}}{6n}. \qquad (7.8)
\end{aligned}
$$

Similarly,

$$
S_4\leq\frac{\sqrt{178}\|g\|_M p}{6n}\leq\frac{\sqrt{178}\|g\|_{M^2}}{6n}. \qquad (7.9)
$$

and, for S_5:

$$
\begin{aligned}
S_5=&\frac{1}{2}\left|\frac{n(n-1)}{8}\mathbb{E}D^2f(\mathbf{Y}_n)\left[(\mathbf{T}_n-\mathbf{T}'_n)(2,0),(\mathbf{V}_n-\mathbf{V}'_n)(0,1)\right]-\frac{4}{3}\mathbb{E}D^2f(\mathbf{Y}_n)\left[\left(\mathbf{D}_n^{(1)},0\right),\left(0,\mathbf{D}_n^{(2,2)}\right)\right]\right|\\
\leq&\frac{\sqrt{178}\|g\|_{M^2}}{12n}. \qquad (7.10)
\end{aligned}
$$

Now, for S_6, let $\mathbf{Y}_n^{ijkl}$ be equal to $\mathbf{Y}_n$ except that $I_{ij},I_{ik},I_{il},I_{jk},I_{jl},I_{kl}$ are replaced with independent copies $I'_{ij},I'_{ik},I'_{il},I'_{jk},I'_{jl},I'_{kl}$, i.e. for all $t\in[0,1]$ let

$$
\begin{aligned}
\mathbf{T}_n^{ijkl}(t)=&\mathbf{T}_n(t)-\frac{\lfloor nt\rfloor-2}{n^2}\left[(I_{ij}-I'_{ij})\mathbb{1}_{[i/n,1]\cap[j/n,1]}(t)+(I_{ik}-I'_{ik})\mathbb{1}_{[i/n,1]\cap[k/n,1]}(t)\right.\\
&+(I_{il}-I'_{il})\mathbb{1}_{[i/n,1]\cap[l/n,1]}(t)+(I_{jk}-I'_{jk})\mathbb{1}_{[j/n,1]\cap[k/n,1]}(t)\\
&\left.+(I_{jl}-I'_{jl})\mathbb{1}_{[j/n,1]\cap[l/n,1]}(t)+(I_{kl}-I'_{kl})\mathbb{1}_{[k/n,1]\cap[l/n,1]}(t)\right]\\
\mathbf{V}_n^{ijkl}(t)=&\mathbf{V}_n(t)-\frac{1}{n^2}\sum_{m:m\neq i,j,k,l}\left[\left(I_{ij}-I'_{ij}\right)\left(I_{im}+I_{jm}\right)\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[m/n,1]}(t)\right.\\
&+\left(I_{ik}-I'_{ik}\right)\left(I_{im}+I_{km}\right)\mathbb{1}_{[i/n,1]\cap[k/n,1]\cap[m/n,1]}(t)\\
&+\left(I_{il}-I'_{il}\right)\left(I_{im}+I_{lm}\right)\mathbb{1}_{[i/n,1]\cap[l/n,1]\cap[m/n,1]}(t)\\
&+\left(I_{jk}-I'_{jk}\right)\left(I_{jm}+I_{km}\right)\mathbb{1}_{[j/n,1]\cap[k/n,1]\cap[m/n,1]}(t)\\
&+\left(I_{jl}-I'_{jl}\right)\left(I_{jm}+I_{lm}\right)\mathbb{1}_{[j/n,1]\cap[l/n,1]\cap[m/n,1]}(t)\\
&\left.+\left(I_{kl}-I'_{ll}\right)\left(I_{km}+I_{lm}\right)\mathbb{1}_{[k/n,1]\cap[l/n,1]\cap[m/n,1]}(t)\right]\\
&-\frac{1}{n^2}\left[(I_{ij}I_{jk}-I'_{ij}I'_{jk})+(I_{ij}I_{ik}-I'_{ij}I'_{ik})+(I_{ik}I_{jk}-I'_{ij}I'_{jk})\right]\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(t)
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{1}{n^2}\left[(I_{ij}I_{jl} - I'_{ij}I'_{jl}) + (I_{ij}I_{il} - I'_{ij}I'_{il}) + (I_{il}I_{jl} - I'_{ij}I'_{jl})\right]\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[l/n,1]}(t)\\
&- \frac{1}{n^2}\left[(I_{ik}I_{kl} - I'_{ik}I'_{kl}) + (I_{ik}I_{il} - I'_{ik}I'_{il}) + (I_{il}I_{kl} - I'_{ik}I'_{kl})\right]\mathbb{1}_{[i/n,1]\cap[k/n,1]\cap[l/n,1]}(t)\\
&- \frac{1}{n^2}\left[(I_{jk}I_{jl} - I'_{jk}I'_{jl}) + (I_{jl}I_{kl} - I'_{jl}I'_{kl}) + (I_{kl}I_{jk} - I'_{kl}I'_{jk})\right]\mathbb{1}_{[j/n,1]\cap[k/n,1]\cap[l/n,1]}(t)
\end{aligned}
\tag{7.11}
$$

and for all $t \in [0,1]$ let $\mathbf{Y}_n^{ijkl}(t) = \left(\mathbf{T}_n^{ijkl}(t) - \mathbb{E}\mathbf{T}_n, \mathbf{V}_n^{ijkl}(t) - \mathbb{E}\mathbf{V}_n(t)\right)$. Note that:

$$
\begin{aligned}
S_6 \leq &\Bigg| \frac{1}{8n^4}\sum_{1\leq i\neq j\leq n}\sum_{k\neq i,j}\sum_{l\neq i,j}\mathbb{E}\left\{(I_{ij}-I'_{ij})^2(I_{jk}+I_{ik})(I_{jl}+I_{il})\right.\\
&\left.\cdot D^2 f(\mathbf{Y}_n)\left[\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(0,1), \mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[l/n,1]}(0,1)\right]\right\}\\
&- \mathbb{E}D^2 f(\mathbf{Y}_n)\left[\sum_{i,j,k=1}^n Z_{i,j,k}^{(2,2)}\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(0,1), \sum_{i,j,k=1}^n Z_{i,j,k}^{(2,2)}\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(0,1)\right]\Bigg|\\
\leq &\Bigg|\sum_{1\leq i\neq j\leq n}\sum_{k\neq i,j}\sum_{l\neq i,j}\mathbb{E}\left\{\left(\frac{1}{8n^4}(I_{ij}-I'_{ij})^2(I_{jk}+I_{ik})(I_{jl}+I_{il}) - Z_{ijk}^{(2,2)}Z_{ijl}^{(2,2)}\right)\right.\\
&\left.\cdot D^2 f(\mathbf{Y}_n)\left[\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(0,1), \mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[l/n,1]}(0,1)\right]\right\}\Bigg|\\
= &\Bigg|\sum_{1\leq i\neq j\leq n}\sum_{k\neq i,j}\sum_{l\neq i,j}\frac{1}{8n^4}\mathbb{E}\left\{(I_{ij}-I'_{ij})^2(I_{jk}+I_{ik})(I_{jl}+I_{il})\right.\\
&\left.\cdot\left(D^2 f(\mathbf{Y}_n) - D^2 f(\mathbf{Y}_n^{ijkl})\right)\left[\mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[k/n,1]}(0,1), \mathbb{1}_{[i/n,1]\cap[j/n,1]\cap[l/n,1]}(0,1)\right]\right\}\Bigg|\\
\leq &\frac{\|g\|_{M^2}}{24n^4}\sum_{\substack{1\leq i,j,k,l\leq n\\ i,j,k,l \text{ distinct}}}\mathbb{E}\left\{(I_{ij} - 2pI_{ij} + p)(I_{jk}+I_{ik})(I_{jl}+I_{il})\|\mathbf{Y}_n - \mathbf{Y}_n^{ijkl}\|\right\}.
\end{aligned}
\tag{7.12}
$$

Now, by (7.11), note that:

$$
\begin{aligned}
&\|\mathbf{Y}_n - \mathbf{Y}_n^{ijkl}\|\\
\leq &\frac{1}{n^2}\Bigg\{(n-2)^2\left(|I_{ij}-I'_{ij}| + |I_{ik}-I'_{ik}| + |I_{il}-I'_{i}| + |I_{jk}-I'_{jk}| + |I_{jl}-I'_{jl}| + |I_{kl}-I'_{kl}|\right)^2\\
&+ \Bigg[\sum_{m:m\neq i,j,k,l}\left[\left|I_{ij}-I'_{ij}\right|(I_{im}+I_{jm}) + \left|I_{ik}-I'_{ik}\right|(I_{im}+I_{km}) + \left|I_{il}-I'_{il}\right|(I_{im}+I_{lm})\right.\\
&\left. + \left|I_{jk}-I'_{jk}\right|(I_{jm}+I_{km}) + \left|I_{jl}-I'_{jl}\right|(I_{jm}+I_{lm}) + \left|I_{kl}-I'_{ll}\right|(I_{km}+I_{lm})\right]
\end{aligned}
$$

$$+|I_{ij}I_{jk}-I'_{ij}I'_{jk}|+|I_{ij}I_{ik}-I'_{ij}I'_{ik}|+|I_{ik}I_{jk}-I'_{ij}I'_{jk}|+|I_{ij}I_{jl}-I'_{ij}I'_{jl}|$$
$$+|I_{ij}I_{il}-I'_{ij}I'_{il}|+|I_{il}I_{jl}-I'_{ij}I'_{jl}|+|I_{ik}I_{kl}-I'_{ik}I'_{kl}|+|I_{ik}I_{il}-I'_{ik}I'_{il}|$$
$$\left.\left.+|I_{il}I_{kl}-I'_{ik}I'_{kl}|+|I_{jk}I_{jl}-I'_{jk}I'_{jl}|+|I_{jl}I_{kl}-I'_{jl}I'_{kl}|+|I_{kl}I_{jk}-I'_{kl}I'_{jk}|\right]^2\right\}^{1/2}$$

$$\le\frac{\sqrt{36(n-2)^2+(12(n-4)+12)^2}}{n^2}$$
$$=\frac{\sqrt{180n^2-1008n+1440}}{n^2}.$$

Therefore, by (7.12):

$$S_7\le\frac{\|g\|_{M^2}\cdot 4\sqrt{180n^2-1008n+1440}}{24n^2}\le\frac{\sqrt{612}\|g\|_{M^2}}{6n}. \tag{7.13}$$

Furthermore, for S_7, note that:

$$S_7=\left|\sum_{i,j=1}^n\mathbb{E}\left(Z_{i,j}^{(2,1)}\right)^2D^2f(\mathbf{Y}_n)\left[\left(0,(\lfloor n\cdot\rfloor-2)\mathbb{1}_{[i/n,1]\cap[j/n,1]}\right),\left(0,(\lfloor n\cdot\rfloor-2)\mathbb{1}_{[i/n,1]\cap[j/n,1]}\right)\right]\right|$$
$$\le\frac{\|g\|_{M^2}}{n}. \tag{7.14}$$

The result now follows by (7.4), (7.5), (7.8), (7.9), (7.10), (7.13), (7.14). □

7.3. *Technical details of the proof of Theorem 6.4*

Lemma 7.4. *Using the notation of* ***Step 1*** *of the proof of Theorem 6.4, for all* $u,t\in[0,1]$,

$$\mathbb{E}\mathbf{D}_n^{(2)}(t)\mathbf{D}_n^{(2)}(u)=\mathbb{E}\mathbf{Z}_n^{(2)}(t)\mathbf{Z}_n^{(2)}(u).$$

Proof. Note that

$$\mathbb{E}\mathbf{D}_n^{(2)}(t)\mathbf{D}_n^{(2)}(u)$$
$$=\mathbb{E}\mathbf{D}_n^{(2,1)}(t)\mathbf{D}_n^{(2,1)}(u)+\mathbb{E}\mathbf{D}_n^{(2,1)}(t)\mathbf{D}_n^{(2,2)}(u)+\mathbb{E}\mathbf{D}_n^{(2,2)}(t)\mathbf{D}_n^{(2,1)}(u)+\mathbb{E}\mathbf{D}_n^{(2,2)}(t)\mathbf{D}_n^{(2,2)}(u)$$
$$=(\lfloor nt\rfloor-2)(\lfloor nu\rfloor-2)\sum_{1\le i,j\le\lfloor n(t\wedge u)\rfloor}\mathbb{E}\left[\left(Z_{ij}^{(2,1)}\right)^2\right]$$
$$+(\lfloor nt\rfloor-2)\sum_{\substack{1\le i,j\le\lfloor n(t\wedge u)\rfloor\\ i\neq j}}\sum_{\substack{1\le k\le\lfloor nu\rfloor\\ k\neq i,k\neq j}}\mathbb{E}Z_{i,j}^{(2,1)}Z_{i,j,k}^{(2,2)}$$
$$+(\lfloor nu\rfloor-2)\sum_{\substack{1\le i,j\le\lfloor n(t\wedge u)\rfloor\\ i\neq j}}\sum_{\substack{1\le k\le\lfloor nt\rfloor\\ k\neq i,k\neq j}}\mathbb{E}Z_{i,j}^{(2,1)}Z_{i,j,k}^{(2,2)}$$

$$
\begin{aligned}
&+\sum_{\substack{1\leq i,j,k\leq\lfloor n(u\wedge t)\rfloor\\ i,j,k \text{ distinct}}}\mathbb{E}\left[\left(Z^{(2,2)}_{i,j,k}\right)^2\right]\\
&+\sum_{\substack{1\leq i,j\leq\lfloor n(u\wedge t)\rfloor\\ 1\leq k\leq\lfloor nu\rfloor,1\leq l\leq\lfloor nt\rfloor\\ i,j,k,l \text{ distinct}}}\mathbb{E}Z^{(2,2)}_{i,j,k}Z^{(2,2)}_{i,j,l}\\
=&\frac{(\lfloor nt\rfloor-2)(\lfloor nu\rfloor-2)\lfloor n(t\wedge u)\rfloor(\lfloor n(t\wedge u)\rfloor-1)}{n^5}\\
&+(\lfloor nt\rfloor-2)(\lfloor nu\rfloor-2)\lfloor n(t\wedge u)\rfloor(\lfloor n(t\wedge u)\rfloor-1)\frac{p^3(1-p)}{n^4}\\
&+\lfloor n(t\wedge u)\rfloor(\lfloor n(t\wedge u)\rfloor-1)(\lfloor n(t\wedge u)\rfloor-2)\frac{p^2(1-p^2)}{2n^4}\\
&+\lfloor n(t\wedge u)\rfloor(\lfloor n(t\wedge u)\rfloor-1)(\lfloor n(u\wedge t)\rfloor-2)(\lfloor n(u\vee t)\rfloor-3)\frac{p^3(1-p)}{n^4}\\
=&(\lfloor nt\rfloor-2)(\lfloor nu\rfloor-2)\lfloor n(t\wedge u)\rfloor(\lfloor n(t\wedge u)\rfloor-1)\frac{2p^3(1-p)}{n^4}\\
&+\lfloor n(t\wedge u)\rfloor(\lfloor n(t\wedge u)\rfloor-1)\\
&\qquad\cdot\left(\frac{(\lfloor nt\rfloor-2)(\lfloor nu\rfloor-2)}{n^5}+\frac{\lfloor n(t\wedge u)\rfloor p^2(1-p)^2}{2n^4}+\frac{2p^3(1-p)}{n^4}\right)\\
=&\mathbb{E}\mathbf{Z}^{(2)}_n(t)\mathbf{Z}^{(2)}_n(u),
\end{aligned}
$$

which finishes the proof. □

Lemma 7.5. *Using the notation of* ***Step 2*** *of the proof of Theorem 6.4,*

$$
\begin{aligned}
&\mathbb{E}\left\|\mathbf{Z}_n-\mathbf{Z}\right\|\leq\frac{12}{n^{1/2}}+\frac{51\sqrt{\log n}}{\sqrt{n}}\\
&\mathbb{E}\left\|\mathbf{Z}_n-\mathbf{Z}\right\|^2\leq\frac{121}{n}+\frac{743\log n}{n}\\
&\mathbb{E}\|\mathbf{Z}\|^2\leq 5.
\end{aligned}
$$

Proof. Note the following

1. By Doob's L^2 inequality,

$$
\begin{aligned}
&\text{A)}\quad\mathbb{E}\left[\sup_{t\in[0,1]}\left|\mathbf{B}_3\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)\right|\right]\leq 2\sqrt{\mathbb{E}\left[\left|\mathbf{B}_3\left(\frac{n(n-1)}{n^2}\right)\right|^2\right]}\leq 2\\
&\text{B)}\quad\mathbb{E}\left[\sup_{t\in[0,1]}\left|\mathbf{B}_4\left(\frac{\lfloor nt\rfloor^2(\lfloor nt\rfloor-1)}{n^3}\right)\right|\right]\leq 2\sqrt{\mathbb{E}\left[\left|\mathbf{B}_4\left(\frac{n^2(n-1)}{n^3}\right)\right|\right]}\leq 2.
\end{aligned}
\tag{7.15}
$$

2. By Doob's L^2 inequality,

$$\mathbb{E}\left[\sup_{t\in[0,1]}|\mathbf{B}_1(t^2)|\right]\leq 2 \quad \text{and} \quad \left|\frac{\lfloor nt\rfloor-2}{n}-t\right|\leq\frac{3}{n} \quad \text{for all } t\in[0,1] \tag{7.16}$$

3. Using [FN10, Lemma 3] and the fact that

$$\left|\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}-t^2\right|\leq\left|\frac{(nt-\lfloor nt\rfloor)(nt+\lfloor nt\rfloor)}{n^2}\right|+\frac{1}{n^2}\leq\frac{3}{n},$$

we obtain

$$\mathbb{E}\left[\sup_{t\in[0,1]}\left|\mathbf{B}_1\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-\mathbf{B}_1(t^2)\right|\right]\leq\frac{30\sqrt{3\log\left(\frac{2n}{3}\right)}}{n^{1/2}\sqrt{\pi\log(2)}}. \tag{7.17}$$

Now, can bound $\mathbb{E}\|\mathbf{Z}_n-\mathbf{Z}\|$ in the following way:

$$\begin{aligned}
&\mathbb{E}\|\mathbf{Z}_n-\mathbf{Z}\|\\
\leq&\frac{\sqrt{p(1-p)}}{\sqrt{2+8p^2}}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_1\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-t\mathbf{B}_1(t^2)\right|\right]\\
&+\frac{p\sqrt{2p(1-p)}}{\sqrt{1+4p^2}}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_2\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-t\mathbf{B}_2(t^2)\right|\right]\\
&+\frac{p\sqrt{2p(1-p)}}{\sqrt{1+4p^2}}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_1\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-t\mathbf{B}_1(t^2)\right|\right]\\
&+\frac{2p^2\sqrt{2p(1-p)}}{\sqrt{1+4p^2}}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_2\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-t\mathbf{B}_2(t^2)\right|\right]\\
&+\frac{1}{n^{1/2}}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_3\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)\right|\right]\\
&+\frac{p(1-p)}{\sqrt{2}n^{1/2}}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\mathbf{B}_4\left(\frac{\lfloor nt\rfloor^2(\lfloor nt\rfloor-1)}{n^3}\right)\right|\right]+\frac{\sqrt{2p^3(1-p)}}{n^2}\mathbb{E}|\mathbf{B}_5(1)|\\
\overset{(7.15)}{\leq}&\frac{(1+4p+4p^2)\sqrt{p(1-p)}}{\sqrt{2+8p^2}}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_1\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-t\mathbf{B}_1(t^2)\right|\right]\\
&+\frac{2}{n^{1/2}}+\frac{\sqrt{2}p(1-p)}{n^{1/2}}+\frac{2\sqrt{p^3(1-p)}}{\sqrt{\pi}n^2}\\
\leq&\frac{(1+4p+4p^2)\sqrt{p(1-p)}}{\sqrt{2+8p^2}}\left(\mathbb{E}\left[\sup_{t\in[0,1]}\left|\left(\frac{\lfloor nt\rfloor-2}{n}-t\right)\mathbf{B}_1(t^2)\right|\right]\right.\\
&\left.+\mathbb{E}\left[\sup_{t\in[0,1]}\left|\mathbf{B}_1\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-\mathbf{B}_1(t^2)\right|\right]\right)+\frac{2+\sqrt{2}p(1-p)}{n^{1/2}}+\frac{2\sqrt{p^3(1-p)}}{\sqrt{\pi}n^2}
\end{aligned}$$

$$\stackrel{(7.16),(7.17)}{\leq} \frac{(1+4p+4p^2)\sqrt{p(1-p)}}{\sqrt{2+8p^2}}\left(\frac{6}{n}+\frac{30\sqrt{3\log n}}{n^{1/2}\sqrt{\pi\log(2)}}\right)+\frac{2+\sqrt{2}p(1-p)}{n^{1/2}}+\frac{2\sqrt{p^3(1-p)}}{\sqrt{\pi}n^2}$$
$$\leq \frac{12}{n^{1/2}}+\frac{51\sqrt{\log n}}{\sqrt{n}}.$$

Similarly, using Doob's L^2 inequality and [FN10, Lemma 3],

$$\begin{aligned}
&\mathbb{E}\|\mathbf{Z}_n-\mathbf{Z}\|^2\\
\leq& 2\frac{p(1-p)}{2+8p^2}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_1\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-t\mathbf{B}_1(t^2)\right|^2\right]\\
&+2\frac{2p^3(1-p)}{1+4p^2}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_2\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-t\mathbf{B}_2(t^2)\right|^2\right]\\
&+5\frac{2p^3(1-p)}{1+4p^2}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_1\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-t\mathbf{B}_1(t^2)\right|^2\right]\\
&+5\frac{8p^5(1-p)}{1+4p^2}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_2\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-t\mathbf{B}_2(t^2)\right|^2\right]\\
&+\frac{5}{n}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_3\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)\right|^2\right]\\
&+\frac{5p^2(1-p)^2}{2n}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\mathbf{B}_4\left(\frac{\lfloor nt\rfloor^2(\lfloor nt\rfloor-1)}{n^3}\right)\right|^2\right]+\frac{2p^3(1-p)}{n^4}\mathbb{E}|\mathbf{B}_5(1)|^2\\
\leq&\frac{p(1-p)(1+14p^2+40p^4)}{1+4p^2}\mathbb{E}\left[\sup_{t\in[0,1]}\left|\frac{\lfloor nt\rfloor-2}{n}\mathbf{B}_1\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-t\mathbf{B}_1(t^2)\right|^2\right]\\
&+\frac{20}{n}+\frac{10p^2(1-p)^2}{n}+\frac{2p^3(1-p)}{n^4}\\
\leq&\frac{p(1-p)(1+14p^2+40p^4)}{1+4p^2}\left(\mathbb{E}\left[\sup_{t\in[0,1]}\left|\left(\frac{\lfloor nt\rfloor-2}{n}-t\right)\mathbf{B}_1(t^2)\right|^2\right]\right.\\
&\left.+\mathbb{E}\left[\sup_{t\in[0,1]}\left|\mathbf{B}_1\left(\frac{\lfloor nt\rfloor(\lfloor nt\rfloor-1)}{n^2}\right)-\mathbf{B}_1(t^2)\right|^2\right]\right)+\frac{20+10p^2(1-p)^2}{n}+\frac{2p^3(1-p)}{n^4}\\
\leq&\frac{p(1-p)(1+14p^2+40p^4)}{1+4p^2}\left(\frac{36}{n^2}+\frac{270\log n}{n\log 2}\right)+\frac{20+10p^2(1-p)^2}{n}+\frac{2p^3(1-p)}{n^4}\\
\leq&\frac{121}{n}+\frac{743\log n}{n}.
\end{aligned}$$

Furthermore, by Doob's L^2 inequality,

$$\mathbb{E}\|\mathbf{Z}\|^2\leq\mathbb{E}\left[\sup_{t\in[0,1]}\left(\frac{\sqrt{p(1-p)}}{\sqrt{2+8p^2}}t\mathbf{B}_1(t^2)+\frac{p\sqrt{2p(1-p)}}{\sqrt{1+4p^2}}t\mathbf{B}_2(t^2)\right)^2\right]$$

$$+\mathbb{E}\left[\sup_{t\in[0,1]}\left(\frac{p\sqrt{2p(1-p)}}{\sqrt{1+4p^2}}t\mathbf{B}_1(t^2)+\frac{2p^2\sqrt{2p(1-p)}}{\sqrt{1+4p^2}}t\mathbf{B}_2(t^2)\right)^2\right]$$

$$\leq\frac{p(1-p)(1+8p^2+16p^4)}{1+4p^2}\mathbb{E}\left[\sup_{t\in[0,1]}|\mathbf{B}_1(t^2)|^2\right]$$

$$\leq\frac{4p(1-p)(1+8p^2+16p^4)}{1+4p^2}\leq 5.$$

This finishes the proof. □

Part III: Summary, conclusions and suggestions for future work

1 Summary

Motivated by the work of Barbour [Bar90], I have approached several problems within the area of Stein's method for infinite-dimensional distributions. In [KDV17], we have corrected a mistake in the construction of a solution to the Stein equation of [Bar90] and confirmed that all the major results of the paper hold true. In [Kas17a] I have concentrated on bounding distances between time-changed Poisson processes and time-changed Wiener process and used the obtained results to present a strategy for bounding the rate of convergence of a certain class of continuous-time Markov chains to diffusions. I have also applied my findings to analyse the asymptotic behaviour of a sequence of rescaled M/M/1 queues. In [Kas18], I have analysed the asymptotics of scaled sums of random vectors under different dependence schemes and provided bounds on their distance from continuous Gaussian processes. The applicability of the bounds obtained in [Kas18] has been presented in concrete examples: a functional central limit theorem for non-degenerate U-statistics and an analysis of the m-scans process. [Kas17b] extends the well-studied exchangeable pair approach to Stein's method to approximations by laws of stochastic processes. Alongside an abstract approximation theorem, it shows its applicability in the proof of a functional combinatorial central limit theorem, strengthening the results of [BJ09], and in the study of edge and two-star counts in a Bernoulli-graph-valued process.

2 Conclusions and discussion

My extension of the results of [Bar90], alongside [CD13, BDM18] and some other papers coauthored by Laurent Decreusefond, make progress at filling in a serious gap in the literature. The error made by practitioners who choose to model real-life discrete phenomena with scaling limits of discrete processes rather than those processes themselves should be quantifiable for the obtained results to be more meaningful. The celebrated strong approximation result [KMT75, KMT76] by Komlós, Major and Tusnády applies only to a limited number of examples which are related to the uniform empirical process. Stein's method may, however, potentially be applied to a much wider class of results due to the fact that it works under very general conditions.

The setup I have considered is the same as that of [Bar90]. I looked at the processes of interest as random elements of the Skorokhod space $D[0,1]$ of càdlàg paths and measured distances between them with respect to test functions acting on this space. Because of its richness, results involving $D[0,1]$ should be of interest to a wide range of practitioners. Considering a restricted Hilbert space, for instance the Besov-Liouville space studied by Decreusefond and coauthors, opens the door for applying Malliavin Calculus and allows one to obtain strong results with very tight bounds. The extent to which such results may be used for real-world applications and be fully understood by applied researchers remains open, though. My intuition is that a real breakthrough will come once a large space with an easy to understand structure is found that could accommodate the Malliavin-Stein approach and allow its use in the functional approximation context.

It is worth noting that my approximations by continuous processes were normally obtained in two steps. The first one involved approximations by a pre-limiting piecewise-constant Gaussian process, which was then compared to the continuous process of interest in the second step. This intuitively does not look like an optimal approach and tools for obtaining a one-step comparison with the continuous process via Stein's method could possibly improve the order of the bounds. The Stein operator for Brownian Motion constructed by Barbour in [Bar90] involves an infinite sum which makes its analysis with the machinery usually used in Stein's method difficult. I welcome the attempts made by several researchers to construct a new Stein operator for Brownian Motion involving Malliavin-calculus-type objects and believe that, if the point of the previous paragraph can be addressed, those could be a source of powerful results.

3 Suggestions for future work

The results of this thesis could be extended and applied in a number of different areas. Some of those are described below.

3.1 Stein's method for the Stroock-Varadhan theory

As mentioned Section 2.2 of Part I of this thesis, the Stroock-Varadhan theory [SV79] provides a class of results about weak convergence of laws of rescaled Markov chains to diffusion processes. It would be interesting to extend the results of [Kas17a], constituting part of this thesis, to cater for a more substantial part of the theory. In particular, applying the decomposition of a continuous-time Markov chain into a sum of time-changed Poisson processes in order to then compare those Poisson processes to time-changed Wiener processes looks like a good approach to follow as long as some technical issues can be overcome.

At the moment, the framework of [Kas17a] is sufficient to treat examples of continuous-time Markov chains which can be decomposed as a sum of Poisson processes time-changed by some (random) time functions independent of those Poisson processes. Providing a framework in which one could prove similar results while adding a complicated dependence structure between the time changes and the processes they are applied to could open the door for a number of very powerful results.

3.2 Degenerate U-statistics and quantitative functional de Jong theorem

The de Jong theorem [dJ90] describes the asymptotic behaviour of degenerate U-statistics. In order to understand the statement, let us set a probability space $(\Sigma, \mathcal{F}, \mathbb{P})$ and for an integer $n \geq 1$, suppose that $X_1, \cdots, X_n$ are independent random variables on this space taking values in measurable spaces $(E_1, \mathcal{E}_1), \cdots, (E_n, \mathcal{E}_n)$. We now follow [DP17] and recall the definitions and the formulation of de Jong's CLT from therein. Let $f : \left(\Pi_{j=1}^n E_j, \bigotimes_{j=1}^n \mathcal{E}_j\right) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable and suppose that

$$Y = f(X_1, \cdots, X_n) \in L^4(\mathbb{P}) \text{ (i.e. has a finite fourth moment)}$$

is centred with unit variance. Let us write

$$Y = \sum_{J \subset [n]} W_J, \tag{10}$$

where, for each J, W_J is $\mathcal{F}_J := \sigma(X_j, j \in J)$-measurable and $\mathbb{E}[W_J|\mathcal{F}_K] = 0$ whenever $J \not\subseteq K$. Representation (10) is called *the Hoeffding decomposition.* It exists and is almost surely unique as long as Y has a finite first moment.

If for some $1 \leq d \leq n$, Y's Hoeffding decomposition takes the form

$$Y = \sum_{J \in \mathcal{D}_d} W_J,$$

where $\mathcal{D}_d = \{J \in [n] : |J| = d\}$, then we call Y a *degenerate U-statistic of order d.* Otherwise, we call it a *non-degenerate U-statistic.*

The celebrated CLT by de Jong states the following. Let $d \geq 1$, $\{n_m : m \geq 1\}$ be a sequence of integers diverging to infinity and $\{W_m\}_{m=1}^{\infty}$ be a sequence of unit-variance degenerate U-statistics of order d, such that each W_m is a function of the vector of independent variables $\left(X_1^{(m)}, \cdots, X_{n_m}^{(m)}\right)$. Suppose that $\mathbb{E}W_m^4 \xrightarrow{m\to\infty} 3$ and $\rho_{n_m}^2 \to 0$, where

$$\rho_n^2 = \max_{1 \leq i \leq n} \sum_{\substack{K \in \mathcal{D}_d \\ i \in K:}} \mathrm{Var}(W_K).$$

Then the law of W_m converges weakly to the standard Gaussian law.

This result is very similar in spirit to a class of results which were proved later and are now known as the *fourth moment theorems.* The most famous ones, provided in [NP05, PT05], state that a sequence of normalised random variables belonging to a fixed Wiener chaos of a Gaussian field converges to the standard Gaussian law in distribution if and only if the sequence of fourth moments of those variables converges to 3, the standard Gaussian fourth moment.

In [DP17], Döbler and Peccati use Stein's method to establish a bound on the rate of convergence in de Jong's theorem and prove its multidimensional version together with the corresponding bounds. In [DP18] they also use contraction operators to prove limit theorems for degenerate and non-degenerate symmetric U-statistics.

In a piece of work in progress, together with Christian Döbler and Giovanni Peccati, we use Stein's method to establish a functional version of de Jong's theorem, in one and multiple dimensions, together with bounds on the rate of convergence. It may be seen as an extension of [Kas18, Theorem 3.3] to the degenerate U-statistic setup. Our techniques are, however, more in the spirit of [Kas17b] as we use the exchangeable-pair approach to Stein's method in the context of functional approximations. Some applications we have been thinking about are connected to testing for qualitative features of functions in non-parametric statistics as outlined in the recent work [CK18].

3.3 Quantitative Breuer-Major theorem

In a recent piece of work [NN18], Nourdin and Nualart prove a functional version of the celebrated Breuer-Major theorem [BM83]. We will follow the setup and notation of [NN18]. The theorem considers a sequence of Gaussian random variables $\{X_n\}_{n\in\mathbb{Z}}$ with mean 0 and covariance function $\mathbb{E}[X_nX_m] = \rho(|n-m|)$ with $\rho(0) = 1$. The classical Breuer-Major theorem states the following. Suppose that

$$\sum_{k\in\mathbb{Z}} |\rho(k)|^d < \infty,$$

and $\varphi \in L^2(\mathbb{R},\gamma)$ (i.e. φ real-valued, such that $\int \varphi^2 d\gamma < \infty$) is of Hermite rank $d \geq 1$ for $\gamma = \mathcal{N}(0,1)$. This means that φ assumes the following expansion

$$\varphi(x) = \sum_{q=d}^{\infty} c_q H_q(x), \quad c_d \neq 0,$$

where $H_q(x)$ is the q-th Hermite polynomial with leading coefficient 1. Then the finite-dimensional distributions of the process

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor - 1} \varphi(X_i), \quad t \in [0,1]$$

converge to those of σW as $n \to \infty$, where $W = \{W_t\}_{t\in[0,1]}$ is a standard Brownian Motion and

$$\sigma^2 = \sum_{q=d}^{\infty} q! c_q^2 \sum_{k\in\mathbb{Z}} \rho(k)^q.$$

Nourdin and Nualart use Malliavin calculus to establish tightness for the process Y_n and so prove the convergence of the entire process in $D[0,1]$ with respect to the Skorokhod topology. It would be interesting to work on bounds on the rate of this convergence. Those were established in [NPP11] for the classical (finite-dimensional) Breuer-Major theorem using the Malliavin-Stein approach of [NP12]. Combining and extending the ideas of [NN18, NPP11] should be an interesting project to look at. I expect the tools used in the project to be very similar to those of [Kas18] as the functional Breuer-Major theorem is a result about a Brownian approximation of a scaled sum of dependent terms.

3.4 Applications to computational statistics and machine learning

Stein's method has recently been used for Monte Carlo diagnostics. Specifically, [GM15] and [GM17] introduce the concept of Stein discrepancy between a Monte Carlo sample and its target distribution and provide methods of bounding it. Furthermore, [CSG16, LLJ16] provide new statistical goodness-of-fit tests constructed using Stein's method, which perform very well when used to evaluate the quality of Monte Carlo samples. In a new piece of work [HM18], Huggins and Mackey introduce random-feature Stein discrepancies. All these methods, however, apply only to finite-dimensional target laws. It would be interesting to determine whether they can be extended to gauge the quality of samples from infinite-dimensional laws, for instance those considered in the celebrated technique of Gaussian Process Regression (see e.g. [RW05]).

The first step towards finding the Stein discrepancy introduced in [GM15], for a target measure μ, is constructing the Stein operator $\mathcal{A}$ for this measure. Indeed, the discrepancy is then defined by:

$$\mathcal{S}(q, \mathcal{A}, \mathcal{G}) := \sup_{f \in \mathcal{G}} \left| \sum_{i=1}^{n} q(x_i) \mathcal{A} f(x_i) \right|,$$

where q is the weight function in the Monte Carlo estimators for expectations under μ, $\{x_1, \cdots, x_n\}$ are the sample points and $\mathcal{G}$ is the domain of $\mathcal{A}$. Bounding the discrepancy for diffusion target laws in order to measure the sample quality would require techniques similar to those used in the project on the Stroock-Varadhan theory. As a next step, it would be interesting to adapt the results of [CSG16, LLJ16] and use the Stein discrepancies to construct statistics with tractable asymptotics, which will be used in goodness-of-fit tests.

Indeed, techniques for Monte Carlo simulation of diffusions have been significantly developed (see, for instance, [HSV09]). Those would provide a class of examples for the potential results to be applied to. It would be particularly interesting to see if Stein's method could improve the existing results on finite-data mean and variance guarantees in Bayesian procedures (for instance Gaussian Process Regression), as introduced in my recent work with Jonathan Huggins, Trevor Campbell and Tamara Broderick [HCKB18, HKCB18]. Using Bayesian methods in practice often requires approximations to the posterior distribution due to a high cost associated with sampling from it. The two papers concentrate on quantifying the quality of such approximations through a newly introduced general version of the distance metric called the (p, ν)-*Fisher norm*. For two distributions η and $\tilde{\eta}$ on a Hilbert space $(\mathbb{H}, \|\cdot\|_{\mathbb{H}})$, some measure ν on $\mathbb{R}$ and a reference measure λ, such that both η and $\tilde{\eta}$ are absolutely continuous with respect to λ, we define the distance in the following way

$$d_{p,\nu}(\eta, \tilde{\eta}) = \left\{ \int \left\| D \log \frac{d\tilde{\eta}}{d\lambda} - D \log \frac{d\eta}{d\lambda} \right\|_{\mathbb{H}}^{p} d\nu \right\}^{1/p},$$

where D is the Frechet derivative. [HCKB18, HKCB18] show that bounds on the (p, ν)-Fisher norms imply Wasserstein bounds and that (p, ν)-Fisher norms are very appealing from the computational point of view. [HCKB18] uses the theory infinite-dimensional SDEs to prove theoretical guarantees on the distances derived from Fisher-like norms between the exact and certain approximate posteriors arising in Gaussian Process Regression. While challenging technically, it might be interesting to use Stein's method for the same problem to see whether it could provide tighter bounds.

Since applied machine learning researchers are interested in Gaussian Processes belonging to Reproducing Kernel Hilbert Spaces, there is scope for the Malliavin Calculus tools of [NP12] to be applied to this kind of settings and used to obtain the bounds. A project of this kind would also be an interesting extension of the results of [Kas17a] to a setup in which the distributions of two non-piecewise-constant processes are compared using Stein's method.

References

[Bar88] A.D. Barbour. Stein's method and Poisson process convergence. *Journal of Applied Probability*, 25:175–184, 1988.

[Bar90] A.D. Barbour. Stein's Method for Diffusion Approximation. *Probability Theory and Related Fields*, 84:297–322, 1990.

[BC05] A.D. Barbour and L.H.Y. Chen, editors. *An Introduction to Stein's Method*. Singapore University Press, 2005.

[BDM18] E. Besançon, L. Decreusefond, and P. Moyal. Stein's method for diffusive limit of Markov processes. arXiv:1805.01691, 2018.

[BHJ92] A.D. Barbour, L. Holst, and S. Janson. *Poisson Approximation*. Oxford Studies in Probability. Clarendon Press, Oxford, 1992.

[BJ09] A.D. Barbour and S. Janson. A functional combinatorial central limit theorem. *Electron. J. of Probab.*, 14(81):2352–2370, 2009.

[BM83] P. Breuer and P. Major. Central limit theorems for non-linear functionals of gaussian fields. *Journal of Multivariate Analysis*, 13(3):425 – 441, 1983.

[CD13] L. Coutin and L. Decreusefond. Stein's method for Brownian Approximations. *Communications on Stochastic Analysis*, 7(3):349–372, 2013.

[CDM05] S. Chatterjee, P. Diaconis, and E. Meckes. Exchangeable pairs and Poisson approximation. *Probab. Surveys*, 2:64–106, 2005.

[CGS11] L.H.Y Chen, L. Goldstein, and Q.-M. Shao. *Normal Approximation by Stein's Method*. Probability and Its Applications. Springer Verlag, Berlin Heidelberg, 2011.

[Che75] L.H.Y. Chen. Poisson approximation for dependent trials. *Ann. Probab.*, 3:534–545, 1975.

[CK18] X. Chen and K. Kato. Jackknife multiplier bootstrap: finite sample approximations to the U-process supremum with applications. arXiv:1708.02705, 2018.

[CM08] S. Chatterjee and E. Meckes. Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 4:257–283, 2008.

[CSG16] K. Chwialkowski, H. Strathmann, and A. Gretton. A Kernel Test of Goodness of Fit. *Proceedings of the 33rd International Conference on Machine Learning*, 48:2606–2615, 2016.

[dJ90] P. de Jong. A central limit theorem for generalized multilinear forms. *Journal of Multivariate Analysis*, 34(2):275 – 289, 1990.

[DP17] C. Döbler and G. Peccati. Quantitative de Jong theorems in any dimension. *Electron. J. Probab.*, 22:35 pp., 2017.

[DP18] C. Döbler and G. Peccati. Limit theorems for symmetric U-statistics using contractions. arXiv:1802.00394, 2018.

[Dur96] R. Durrett. *Stochastic Calculus: A Practical Introduction.* CRC Press, Boca Raton, 1996.

[EK86] S.N. Ethier and T.G. Kurtz. *Markov processes: characterization and convergence.* Wiley, New York, 1986.

[GM15] J. Gorham and L. Mackey. Measuring sample quality with Stein's method. *Advances in Neural Information Processing Systems (NIPS)*, Dec 2015.

[GM17] J. Gorham and L. Mackey. Measuring sample quality with kernels. *International Conference on Machine Learning (ICML)*, Aug 2017.

[Göt91] F. Götze. On the rate of convergence in the multivariate CLT. *Ann. Probab.*, 19(2):724–739, 1991.

[HCKB18] J.H. Huggins, T. Campbell, M.J. Kasprzak, and T. Broderick. Scalable Gaussian Process Inference with Finite-data Mean and Variance Guarantees. arXiv:1806.10234, 2018.

[HKCB18] J.H. Huggins, M.J. Kasprzak, T. Campbell, and T. Broderick. Practical bounds on the error of Bayesian posterior approximations: A nonasymptotic approach. arXiv:1809.09505, 2018.

[HM18] J.H. Huggins and L. Mackey. Random Feature Stein Discrepancies. arXiv:1806.07788, 2018.

[HSV09] M. Hairer, A. Stewart, and J. Voss. Sampling conditioned diffusions. In *LMS Lecture Note Series*, volume 353, pages 159–186. 2009.

[Kas17a] M.J. Kasprzak. Diffusion approximations via Stein's method and time changes. arXiv:1701.07633, 2017.

[Kas17b] M.J. Kasprzak. Multivariate functional approximations with Stein's method of exchangeable pairs. arXiv:1710.09263, 2017.

[Kas18] M.J. Kasprzak. Stein's method for multivariate Brownian approximations of sums under dependence. arXiv:1708.02521, 2018.

[KDV17] M.J. Kasprzak, A. B. Duncan, and S.J. Vollmer. Note on A. Barbour's paper on Stein's method for diffusion approximations. *Electron. Commun. Probab.*, 22(23):1–8, 2017.

[KMT75] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV'-s, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32(1):111–131, Mar 1975.

[KMT76] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheor. Verw. Geb*, 34(1):33–58, Mar 1976.

[LLJ16] Q. Liu, J. Lee, and M. Jordan. A kernelized Stein discrepancy for goodness-of-fit tests. *Proceedings of the 33rd International Conference on Machine Learning*, 48:276–284, 2016.

[LRS17] Ch.e Ley, G. Reinert, and Y. Swan. Stein's method for comparison of univariate distributions. *Probab. Surveys*, 14:1–52, 2017.

[Luk94] H.M. Luk. *Stein's method for the gamma distribution and related statistical applications*. PhD thesis, University of Southern California, 1994.

[Mec09] E. Meckes. *On Stein's method for multivariate normal approximation*, volume Volume 5 of *Collections*, pages 153–178. Institute of Mathematical Statistics, Beachwood, Ohio, USA, 2009.

[MRS18] G. Mijoule, G. Reinert, and Y. Swan. Stein operators, kernels and discrepancies for multivariate continuous distributions. arXiv:1806.03478, 2018.

[NN18] I. Nourdin and D. Nualart. The functional Breuer-Major theorem. arXiv:1808.02378, 2018.

[NP05] D. Nualart and G. Peccati. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, 33(1):177–193, 01 2005.

[NP12] I. Nourdin and G. Peccati. *Normal Approximations with Malliavin Calculus*. Cambridge tracts in Mathematics. Cambridge University Press, 2012.

[NPP11] I. Nourdin, G. Peccati, and M. Podolskij. Quantitative Breuer–Major theorems. *Stochastic Processes and their Applications*, 121(4):793 – 812, 2011.

[PT05] G. Peccati and C. A. Tudor. *Gaussian Limits for Vector-valued Multiple Stochastic Integrals*, pages 247–262. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[Rob03] P. Robert. *Stochastic Networks and Queues*. Springer-Verlag Berlin Heidelberg, 2003.

[Röl07] A. Röllin. Translated poisson approximation using exchangeable pair couplings. *Ann. Appl. Probab.*, 17(5/6):1596–1614, 10 2007.

[Ros11] N. Ross. Fundamentals of Stein's Method. *Probability Surveys*, 8:210–293, 2011.

[RR97] Y. Rinott and V. Rotar. On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics. *Ann. Appl. Probab.*, 7(4):1080–1105, 11 1997.

[RR09] G. Reinert and A. Röllin. Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.*, 37(6):2150–2173, 2009.

[RR10] G. Reinert and A. Röllin. Random subgraph counts and u-statistics: Multivariate normal approximation via exchangeable pairs and embedding. *Journal of Applied Probability*, 47(2):378–393, 2010.

[RW05] C. E. Rasmussen and Ch. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning).* The MIT Press, 2005.

[SKM93] S.G. Samko, A.A. Kilbas, and O. I. Marichev. *Fractional integrals and derivatives: theory and applications.* Gordon and Breach Science Publishers, Philadelphia, PA., USA, 1993.

[Ste72] Ch. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.*, 2:583–602, 1972.

[Ste86] Ch. Stein. *Approximate Computation of Expectations.* Institute of Mathematical Statistics Lecture Notes, Monograph Series, 7. Hayward, Calif. Institute of Mathematical Statistics, 1986.

[SV79] D.W. Stroock and S.R.S. Varadhan. *Multidimensional diffusion processes.* Springer-Verlag, Berlin ; New York, 1979.

[Swa16] Y. Swan. A gateway to Stein's Method. https://sites.google.com/site/steinsmethod/home, 2016. Accessed on 19/05/2016.