New developments in Stein’s method
with applications

Dissertation
vorgelegt von
Christian Döbler

Betreuer: Prof. Dr. Peter Eichelsbacher

Fakultät für Mathematik
der Ruhr-Universität Bochum

26. September 2012
Preface

Stein's method is a popular tool for proving distributional convergence of a sequence of random variables. One of its main advantages over other techniques is that it usually gives concrete error bounds on the distance between the involved distributions in several metrics and, hence, automatically yields rates of convergence. Being first developed for the univariate standard normal distribution in [Ste72], [Ste86] and within much subsequent work, it was quickly recognized that Stein's technique of a characterization through an appropriate differential equation is by no means restricted to the normal distribution. In fact, already in [Che75] the method was used to bound the distance to the Poisson distribution for a sum of dependent indicator random variables by replacing the differential equation by a suitable difference equation. When applying Stein's method to a given random variable \( W \), whose distribution is supposed to be close to a given distribution \( \mu \), whose Stein characterization is at hand, one usually needs one more tool, which might for instance be a certain coupling \((W, W')\) for \( W \) or a Stein characterization for \( \mathcal{L}(W) \), the distribution of \( W \). One such classical coupling is the exchangeable pairs coupling, whose usefulness for univariate normal approximation was highlighted in the monograph [Ste86] and which was further extended in many following articles (see, e.g. [RR97], [SS06] and [Röl08]). In the recent papers [EL10] and [CS11], the exchangeable pairs approach for a given univariate, absolutely continuous distribution \( \mu \) was developed in the context of the so-called density approach, which is a universal method of finding a Stein characterization for such distributions \( \mu \). In [Ste86] and [RR97] it was shown that the exchangeable pairs approach for approximation by the standard normal distribution may be successfully applied given the exchangeable pair \((W, W')\) satisfies the linear regression property \( E[W' - W | W] = -\lambda W + R \), where \( \lambda > 0 \) is constant and \( R \) is a negligible remainder term. The question, what regression property to demand of an exchangeable pair for a general distribution \( \mu \) within the density approach was independently answered in [EL10] and [CS11]. They pointed out that the right condition is \( E[W' - W | W] = -\lambda \psi(W) + R \), where \( \lambda \) and \( R \) are as before and where \( \psi \) is the logarithmic derivative of the density of \( \mu \).

Having reviewed Stein's method for the standard normal distribution and having given an abstract account of Stein's method in Chapter 1, we turn to absolutely continuous univariate distributions in Chapter 2. After presenting several approaches for finding a Stein characterization we propose a new such approach,
which is motivated by a regression property, which is satisfied by a given exchangeable pair, whose members are supposed to be close to the given distribution $\mu$, see Section 2.4. The elaboration of this approach is the theoretical centerpiece of this thesis. Afterwards, the theory from Section 2.4 is specialised to the family of Beta distributions in Section 2.5 and is then applied to Pólya urn models, Wigner’s semi-circle law and the arcsine law in Sections 2.6, 2.7 and 2.8.

In Chapter 3 we turn to multivariate normal approximation by Stein’s method, as developed in [Göt91], [Bar90], [CM08], [RR09], [Mec09] and further articles. Having reviewed and extended the exchangeable pairs approach for multivariate normal approximation in Section 3.1, we turn to spectral properties of random Haar distributed elements from one of the classical compact, connected Lie groups in Section 3.2. There, we prove a quantitative version of a theorem byDiaconis and Shahshahani on asymptotic normality and independence of the vector of traces of various powers of a Haar distributed element from one of the classical compact Lie groups, see [DS94], and apply it to obtain rates of convergence for the Gaussian fluctuations of suitable linear statistics of the spectral measure of a Haar distributed unitary matrix.

There are several people, who I am indebted to. First of all, I would like to thank my family and particularly my parents, Karin Döbler and Manfred Döbler, for their constant and extraordinary support and plenty of encouragement.

Then, I thank Peter Eichelsbacher, the advisor of this thesis, as well as Michael Stolz for our collaboration and for drawing my interest on several fruitful topics, which I would not have dealt with otherwise.

I would also like to thank the other former and current members of our working group and especially Hanna Döring, Bastian Martschink, Anselm Reichenbachs and Kai Krokowski for a lot of helpful and also amusing conversation as well as for the cordial atmosphere.

I would like to thank Professors Andrew Barbour and Louis Chen for inviting me to participate in the conference Progress in Stein’s method held at the National University of Singapore in 2009.

I thank the Deutsche Forschungsgemeinschaft for supporting me via SFB TR/12.
Contents

1. **Fundamentals of Stein’s method** ..................................................... 7
   1.1. Stein’s method for univariate normal approximation ................. 8
   1.2. Abstract view on Stein’s method ........................................... 10

2. **Stein’s method for absolutely continuous univariate distributions** .... 15
   2.1. Principles of finding a Stein type characterization ................. 16
       2.1.1. The differential equation approach .................................. 16
       2.1.2. The density approach .............................................. 19
       2.1.3. The parametric approach by Ley and Swan ....................... 22
       2.1.4. The Malliavin calculus approach .................................. 22
   2.2. The problem of support ..................................................... 23
   2.3. The exchangeable pairs approach for absolutely continuous univari-
        ate distributions .......................................................... 25
       2.3.1. Exchangeable pairs and the density approach ................... 26
   2.4. Modification of Stein’s equation by exchangeable pairs ............. 27
   2.5. Stein’s method for Beta distributions .................................. 63
   2.6. The Pólya urn model ....................................................... 77
   2.7. Wigner’s semicircle law .................................................... 82
   2.8. The arcsine law .............................................................. 88

3. **Multivariate normal approximation and the classical compact groups** 97
   3.1. Stein’s method for multivariate normal distributions ............... 98
       3.1.1. The Stein equation and its solution ............................... 98
       3.1.2. The exchangeable pairs approach for multivariate normal
              approximation ..................................................... 111
   3.2. Spectral properties of large random matrices from the classical com-
        pact groups ................................................................... 125
       3.2.1. Traces of powers ...................................................... 127
       3.2.2. Linear eigenvalue statistics ......................................... 156

A. **Appendix** .............................................................................. 165
   A.1. A general version of de l’Hôpital’s rule ................................. 166
   A.2. The Gibbs sampler ............................................................ 172
1. Fundamentals of Stein’s method

As the title already suggests, this chapter aims at introducing Stein’s method of distributional approximation. In Section 1.1 we begin with a short account on the method for univariate normal approximation, to which the method was first applied by Charles Stein in [Ste72]. In particular, we introduce the concepts of a Stein characterization, a Stein equation and the corresponding Stein solution, which are essential ingredients of Stein’s method. Furthermore, we will state a typical lemma, which gives important bounds on the Stein solutions for various test functions.

Then, in Section 1.2 we generalize the concepts from Section 1.1 to give an abstract notion of Stein’s method for distributions on general spaces. Having defined the notion of a Stein operator, we treat the problem of finding such a characterizing operator for a given distribution by presenting the so-called generator approach, which was suggested by Barbour in [Bar88] and which yields both, a Stein operator and a Stein solution to a corresponding Stein equation. Finally, following the work [Göt91] of Götze, we show by example of the multivariate normal approximation how this technique may be applied.
1. Fundamentals of Stein’s method

1.1. Stein’s method for univariate normal approximation

In this motivational section we will review the concrete case of univariate normal approximation by Stein’s method, which goes back to Charles Stein’s seminal paper [Ste72] and was further developed in his monograph [Ste86]. Stein’s method began with the following observation. Let \( \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \) be the continuous density of the standard normal distribution.

**Proposition 1.1.1** (Stein’s characterization of \( \mathcal{N}(0,1) \)). A real-valued random variable \( X \) has the univariate standard normal distribution if and only if for all absolutely continuous functions \( f : \mathbb{R} \to \mathbb{R} \) with \( \int_{\mathbb{R}} |f'(x)| \varphi(x) dx < \infty \) the expected values \( E[f'(X)] \) and \( E[Xf(X)] \) exist and coincide.

Having this characterization at hand, for a given \( \mathcal{N}(0,1) \)-integrable test function \( h : \mathbb{R} \to \mathbb{R} \), it then was Stein’s idea to consider the following differential equation, which is now called Stein’s equation.

\[
f'(x) - xf(x) = h(x) - E[h(Z)] \tag{1.1}
\]

Here, we let \( Z \) have distribution \( \mathcal{N}(0,1) \). If \( f_h \) is a solution to (1.1), then for a given real-valued random variable \( W \), taking expectations on both sides yields

\[
E[f'_h(W) - Wf_h(W)] = E[h(W)] - E[h(Z)]. \tag{1.2}
\]

If \( W \) is supposed to have a distribution which is close to \( \mathcal{N}(0,1) \) in some respect, then one often aims at bounding above a probabilistic metric of the form

\[
d_H(L(W), \mathcal{N}(0,1)) := \sup_{h \in \mathcal{H}} |E[h(W)] - E[h(Z)]|,
\]

where \( \mathcal{H} \) is some class of test functions \( h \) on \( \mathbb{R} \). By equation (1.2) this last supremum also equals the quantity

\[
\sup_{h \in \mathcal{H}} \left| E[f'_h(W) - Wf_h(W)] \right|
\]

and the curious fact is that this quantity is often easier to bound than the original object. This is mainly due to the fact, that in contrast to the right hand side the left hand side of (1.2) only contains one distribution, namely the one of \( W \). Since one can prove uniform bounds on suitable solutions \( f_h \) to (1.1) and on their derivatives for various classes \( \mathcal{H} \), one may use coupling techniques on \( W \) and Taylor expansion to obtain bounds on the given metric. For the normal distribution a lot is known. It turns out, that a solution to (1.1) is given by
1.1. Stein's method for univariate normal approximation

\[ f_h(x) := \frac{1}{\varphi(x)} \int_{-\infty}^{x} (h(y) - E[h(Z)]) \varphi(y) dy = -\frac{1}{\varphi(x)} \int_{x}^{\infty} (h(y) - E[h(Z)]) \varphi(y) dy \]

(1.3)

and that this solution has the best boundedness properties among all solutions, since the solutions of the corresponding homogeneous equation are the constant multiples of \( e^{-x^2/2} \). In order to give a picture of what can be achieved for the solutions \( f_h \), we state the following result without proof.

**Proposition 1.1.2.** Let \( h : \mathbb{R} \to \mathbb{R} \) be a Borel-measurable function with \( E[|h(Z)|] < \infty \) and let \( f_h \) be given by (1.3).

(a) If \( h \) is bounded, then

\[ \|f_h\|_\infty \leq \sqrt{\pi/2} \|h - E[h(Z)]\|_\infty \quad \text{and} \quad \|f'_h\|_\infty \leq 2\|h - E[h(Z)]\|_\infty . \]

(b) If \( h \) is Lipschitz-continuous with minimal Lipschitz constant \( \|h'\|_\infty \), then

\[ \|f_h\|_\infty \leq \|h'\|_\infty , \quad \|f'_h\|_\infty \leq \sqrt{2/\pi} \|h'\|_\infty \quad \text{and} \quad \|f''_h\|_\infty \leq 2\|h'\|_\infty . \]

A proof of Proposition 1.1.2 can be found in the book [CGS11], which gives a comprehensive introduction to Stein's method for normal approximation (but for the factor \( 1 \) in the first bound in the second line see Remark 2.4.22). From Proposition 1.1.2 we see that, in general, \( f_h \) is one order smoother than \( h \). Also note, that these bounds do in no way depend on a special approximation problem at hand, but might be used in a universal way, whenever one is trying to compute a bound on the distance to normality for the distribution of a given random variable \( W \).

In practice, given such a \( W \), one usually needs some further tool to bound the expression

\[ \left| E[f'_h(W) - Wf_h(W)] \right|. \]

This tool can be for example some coupling \( (W, W') \), where \( W' \) is another random variable defined on the same space as \( W \), or a Stein characterization for the distribution \( L(W) \), similar to Proposition 1.1.1 for the standard normal distribution. Prominent coupling constructions in the context of normal approximation are the exchangeable pair coupling by Stein [Ste86], the size-bias coupling by Goldstein and Rinott [GR96] and the zero-bias coupling by Goldstein and Reinert [GR97]. All these couplings and their role in normal approximation by Stein's method are extensively discussed in [CGS11].
1. Fundamentals of Stein’s method

1.2. Abstract view on Stein’s method

The aim of this section is to abstract from the concrete setting of normal approximation and illustrate what Stein’s method for an arbitrary distribution \( \mu \) on a general measurable space \((\mathcal{X}, \mathcal{B})\) is about. To this end, it might be helpful to review Stein’s method for the univariate normal distribution by introducing the differential operator \( L_N \), defined by

\[
L_N f(x) := (L_N f)(x) := f'(x) - xf(x)
\]  

and that Stein’s characterization of \( N(0, 1) \), Proposition 1.1.1, might be paraphrased as that \( X \) has distribution \( N(0, 1) \) if and only if \( E[L_N f(X)] = 0 \) for all sufficiently smooth functions \( f : \mathbb{R} \to \mathbb{R} \). Then, Stein’s equation (1.1) becomes

\[
L_N f(x) = h - E[h(Z)].
\]  

There might be some hope, that for our given distribution \( \mu \) on \((\mathcal{X}, \mathcal{B})\) there is also some operator \( L \) defined on some space \( \text{dom}(L) \) of real-valued functions on \( \mathcal{X} \) such that a given \( \mathcal{X} \)-valued random variable \( X \) has distribution \( \mu \) if and only if

\[
E[L f(X)] = 0
\]  

for all \( f \in \text{dom}(L) \). In this case \( L \) will be called a Stein operator for \( \mu \). Having this characterization for \( \mu \), for a given \( \mu \)-integrable test function \( h : \mathcal{X} \to \mathbb{R} \) one is again led to consider the corresponding Stein equation, which reads

\[
L f = h - \mu(h),
\]  

where we write \( \mu(h) \) for \( \int_{\mathcal{X}} hd\mu \).

Of course, the decisive question is how to find a suitable Stein operator \( L \) such that the Stein identity (1.6) is valid. For absolutely continuous distributions on \((\mathbb{R}, \mathcal{B})\) there already exist several approaches, some of which will be presented in Chapter 2. There, we will also present a new method of finding such a characterizing operator, if a suitable exchangeable pair is given.

The only general procedure to find a Stein type characterization, which we would like to present, is the so-called generator approach proposed by Barbour in [Bar88], who applied it to multivariate Poisson approximation. This approach was also successfully used by Götze in [Göt91] for the development of Stein’s method for the multivariate standard normal distribution and by Barbour for diffusion approximation in [Bar90]. A more detailed account of Götze’s method is given in the paper [BH10].
1.2. Abstract view on Stein’s method

To find a Stein operator for the distribution $\mu$ on the space $(\mathcal{X}, \mathcal{B})$ by the generator approach according to [Bar88], one seeks an ergodic, time-continuous Markov process $(X_t)_{t \geq 0}$, which has invariant distribution $\mu$. We denote by $(T_t)_{t \geq 0}$ the operator semigroup corresponding to $(X_t)_{t \geq 0}$, which acts on $L^2(\mu) := L^2(\mathcal{X}, \mathcal{B}, \mu)$ by

$$T_t f(x) := E\left[f(X_t) \mid X_0 = x\right]. \quad (1.8)$$

The infinitesimal generator $L$ corresponding to the process $(X_t)_{t \geq 0}$ or to the semigroup $(T_t)_{t \geq 0}$ is given by the following limit in $L^2(\mu)$

$$L f := \lim_{t \downarrow 0} \frac{T_t f - f}{t}, \quad (1.9)$$

for all $f \in \text{dom}(L)$, the class of $L^2(\mu)$-functions for which the limit exists. From the general theory of operator semigroups on Banach spaces (see, e.g. [EK86]), one has the following identities for each $t > 0$:

1. $\int_0^t T_s f ds \in \text{dom}(L)$ for all $f \in L^2(\mu)$ and $L\left(\int_0^t T_s f ds\right) = T_t f - f$
2. $T_t(\text{dom}(L)) \subseteq \text{dom}(L)$
3. $\frac{d}{dt}T_t f = LT_t f = T_t Lf$ for all $f \in \text{dom}(L)$

Here, the integral in (1) is an $L^2(\mu)$-valued Riemann integral. Having introduced the necessary concepts, we can now state the characterization result for the generator approach.

**Proposition 1.2.1.** An $(\mathcal{X}, \mathcal{B})$-valued random variable $X$ has distribution $\mu$ if and only if for all functions $f \in \text{dom}(L)$ it holds that $E[Lf(X)] = 0$. Thus, $L$ is a Stein operator for $\mu$.

**Proof.** First, let $X$ have distribution $\mu$ and let $f \in \text{dom}(L)$ be arbitrary. Since $\mu$ is an invariant measure for $(X_t)_{t \geq 0}$ we have

$$\int_{\mathcal{X}} T_t f d\mu = E[T_t f(X)] = E_{\mu}[f(X_t)] = E_{\mu}[f(X_0)] = E[f(X)] = \int_{\mathcal{X}} f d\mu$$

or, equivalently,

$$0 = \int_{\mathcal{X}} \frac{T_t f - f}{t} d\mu$$

for each $t > 0$. Since the integrand here converges in $L^2(\mu)$ to $Lf$ and since the constant $1 \in L^2(\mu)$ we conclude
1. Fundamentals of Stein’s method

\[ 0 = \lim_{t \downarrow 0} \int_{\mathcal{X}} \frac{T_t f - f}{t} d\mu = \int_{\mathcal{X}} \lim_{t \downarrow 0} \frac{T_t f - f}{t} d\mu = \int_{\mathcal{X}} Lf d\mu = E[Lf(X)], \]

as desired.

For the converse let \( f \in L^2(\mu) \) and \( t > 0 \) be given and note that by (1) we can write

\[ E[T_t f(X)] - E[f(X)] = E[T_t f(X) - f(X)] = E[L\left(\int_0^t T_s f ds\right)(X)] = 0, \]

since the function \( \int_0^t T_s f ds \) belongs to \( \text{dom}(L) \) by (1). This proves that \( \mathcal{L}(X) \) is an invariant distribution for \((X_t)_{t \geq 0}\) and by ergodicity of the process it follows that \( \mathcal{L}(X) = \mu \).

In order to solve Stein’s equation (1.7) within the generator approach, we again appeal to identity (1) but with \( f \) replaced by the centered test function \( h - \mu(h) \):

\[ L\left(-\int_0^t (T_s h - \mu(h))ds\right) = h - T_t h \quad (1.10) \]

Letting formally \( t \to \infty \) we arrive at the identity

\[ L\left(-\int_0^\infty (T_s h - \mu(h))ds\right) = h - \mu(h) \quad (1.11) \]

since \((X_t)_{t \geq 0}\) is an ergodic process with invariant distribution \( \mu \). Thus, in general, the generator approach automatically yields both, a Stein operator \( L \) given by (1.9) and a solution \( f_h \) to the corresponding Stein equation (1.7) given by

\[ f_h := -\int_0^\infty (T_s h - \mu(h))ds. \quad (1.12) \]

It is known (see, e.g. [BH10]) that there are general conditions on the semigroup \((T_t)_{t \geq 0}\) or, equivalently, on the infinitesimal generator \( L \) that justify the formal calculations involving the limit \( t \to \infty \). For example, if \((T_t)_{t \geq 0}\) satisfies the spectral gap inequality

\[ \int_{\mathcal{X}} \left(T_t f(x) - \mu(f)\right)^2 d\mu(x) \leq e^{-2\lambda t} \int_{\mathcal{X}} \left(f(x) - \mu(f)\right)^2 d\mu(x), \quad (1.13) \]

for all \( f \in L^2(\mu) \), where \( \lambda > 0 \) is a fixed constant, then the formal computation can be justified. It is known (see e.g. [ABC+00], Chapter 2) that the validity of the
spectral gap inequality (1.13) is in fact equivalent to a Poincaré inequality with constant $c = 1/\lambda$. This means that the measure $\mu$ satisfies
\[
\int_{\mathcal{X}} \left( f(x) - \mu(f) \right)^2 \, d\mu(x) \leq -c \int_{\mathcal{X}} f(x) L f(x) \, d\mu(x) \quad (1.14)
\]
for all $f$ in a suitable function algebra, which is dense in $L^2(\mu)$. If one knows the distribution of $X_s$ for each $s > 0$ given that the process starts deterministically in $x \in \mathcal{X}$, then one may explicitly compute

\[
T_s h(x) = E\left[ h(X_s) \mid X_0 = x \right]
\]
and, hence, by (1.12) also $f$ can be computed. In the case where $\mu$ is the standard normal distribution on $\mathbb{R}^d$, one may take for $(X_t)_{t \geq 0}$ the $d$-dimensional Ornstein-Uhlenbeck process, which is defined by the stochastic differential equation
\[
dX_t = \sqrt{2} dB_t - X_t \, dt, \quad X_0 = x, \quad (1.15)
\]
where $(B_t)_{t \geq 0}$ denotes $d$-dimensional standard Brownian motion. The infinitesimal generator $L$ of the corresponding Ornstein-Uhlenbeck semigroup is given by
\[
L f(x) = \Delta f(x) - \langle x, \nabla f(x) \rangle, \quad (1.16)
\]
where $\Delta$ is the Laplace operator on $\mathbb{R}^d$, $\nabla$ is the gradient operator and $\langle \cdot, \cdot \rangle$ denotes the standard inner product on $\mathbb{R}^d$. There exists a strong solution to (1.15), which is given by
\[
X_t = e^{-t} x + \sqrt{2} e^{-t} \int_0^t e^s \, dB_s. \quad (1.17)
\]
Since $e^s$ is a deterministic function, it is well-known that the Itô integral in (1.17) is normally distributed and in fact that for each $t > 0$ one has
\[
X_t \overset{D}{=} e^{-t} x + \sqrt{1 - e^{-2t}} Z,
\]
where $Z$ has the $d$-dimensional standard normal distribution. This was used by Götze [Göt91], Barbour [Bar90] and others to cope with the solutions $f_h$ of (1.16) in this special case. By the change of variables $t = e^{-2s}$ from (1.12) we obtain
\[
f_h(x) = -\int_0^\infty \left( E\left[ h(e^{-s} x + \sqrt{1 - e^{-2s}} Z) \right] - E[h(Z)] \right) ds
\]
\[
= -\int_0^1 \frac{1}{2t} \left( E\left[ h(\sqrt{t} x + \sqrt{1-t} Z) \right] - E[h(Z)] \right) dt. \quad (1.18)
\]
1. Fundamentals of Stein’s method

Formula (1.18) may now be used to prove bounds on the derivatives of \( f_h \), comparable to those in Proposition 1.1.2 for the univariate normal distribution. For more details see Section 3.1.
2. Stein’s method for absolutely continuous univariate distributions

In this chapter we deal with Stein’s method for univariate distributions, which have a density with respect to Borel-Lebesgue measure. In Section 2.1 we first review some parts of the existing theory on Stein’s method for absolutely continuous univariate distributions, discuss different approaches and reveal connections between them. Then, in Section 2.2 we discuss the problem that the distribution of the given random quantity $W$ need not have the same support as the target distribution, to which it shall be compared. In Section 2.3 we introduce the exchangeable pairs approach for absolutely continuous univariate distributions and present existing theory within the framework of the so-called density approach. Section 2.4 suggests a method of finding a Stein type characterization, which is motivated by and adapted to a certain regression property of a given exchangeable pair, whose members are supposed to be close in distribution to the distribution considered. Within this abstract section, which is the theoretical centerpiece of the present chapter, we also develop a suitable exchangeable pairs approach. This quite general theory is then specialized to the family of Beta distributions in Section 2.5 and Sections 2.6, 2.7 and 2.8 provide concrete applications of Stein’s method for Beta distributions from Section 2.5.
2. Stein’s method for absolutely continuous univariate distributions

2.1. Principles of finding a Stein type characterization

In the whole section let \( \mu \) be an absolutely continuous probability distribution on \((\mathbb{R}, \mathcal{B})\) with density \( p \) and let \( Z \sim \mu \). Throughout, let \(-\infty \leq a < b \leq \infty\) be such that \( \text{supp}(\mu) \subseteq (a, b) \) and for any \( a' > a \) and \( b' < b \) we have \( \text{supp}(\mu) \not\subseteq (a', b') \), i.e. if \( a \) is a real number it is the left endpoint of \( \text{supp}(\mu) \) and if \( b \) is a real number it is the right endpoint of \( \text{supp}(\mu) \). Later we will focus on the case that \( \text{supp}(\mu) = (a, b) \) is in fact an interval.

In the following we will explain how to find a Stein identity and a Stein characterization for \( \mu \) from scratch.

2.1.1. The differential equation approach

Suppose that the density \( p \) is sufficiently smooth on \((a, b)\) and that there is some differential operator \( A \) such that \( Ap = 0 \). Letting \( A^* \) be the formal adjoint of \( A \) in \( L^2((a, b), \lambda) \) we obtain for \( g \in \text{dom}(A^*) \) and \( Z \sim \mu \)

\[
E[A^*g(Z)] = \int_{(a,b)} A^*g(x)p(x)dx = \int_{(a,b)} g(x)Ap(x)dx = \int_{(a,b)} g(x) \cdot 0 dx = 0.
\]

So we have found a Stein identity for \( \mu \). Moreover, at least if the image of \( A^* \) is rich enough, \( A^* \) characterizes \( \mu \). In any case, the Stein equation corresponding to the \( \mu \)-integrable test function \( h \) becomes

\[
A^*g = h - \mu(h) = h - E[h(Z)].
\]

EXAMPLE 2.1.1. (a) The standard normal distribution \( N(0, 1) \)

In this case \( (a, b) = \mathbb{R} \) and \( p(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \). Thus

\[
p'(x) = -xp(x) \quad \text{or} \quad 0 = Ap(x) := p'(x) + xp(x).
\]

Let us compute the adjoint operator \( A^* \) through integration by parts: For \( f, g \) being in \( \mathcal{S}(\mathbb{R}) \), the class of Schwartz functions or rapidly decreasing functions,

\[
\int_{\mathbb{R}} Af(x)g(x)dx = \int_{\mathbb{R}} (f'(x) + xf(x))g(x)dx = \int_{-\infty} f'g \big|_{-\infty}^\infty - \int_{\mathbb{R}} f(x)g'(x)dx + \int_{\mathbb{R}} xf(x)g(x)dx
\]

\[
= \int_{\mathbb{R}} f(x)(-g'(x) + xg(x))dx.
\]
2.1. Principles of finding a Stein type characterization

Hence, \( A^*g(x) = -g'(x) + xg(x) \). Note that \(-A^* \) is the well known Stein operator for \( N(0,1) \).

(b) The semi-circle law on \([-1,1] \)

In this case \((a,b) = [-1,1] \), \( p(x) = \frac{2}{\pi} \sqrt{1-x^2} \) \(1_{[-1,1]}(x) \) and we have for \(-1 < x < 1 \)

\[
p'(x) = -\frac{2}{\pi} \frac{x}{\sqrt{1-x^2}} \quad \text{or} \quad (1-x^2)p'(x) + xp(x) =: Ap(x) = 0.
\]

Again, we use integration by parts to compute the adjoint operator \( A^* \):

\[
\int_{-1}^{1} Af(x)g(x)dx = \int_{-1}^{1} \left( (1-x^2)f'(x) + xf(x) \right) g(x)dx \\
= (1-x^2)f(x)g(x)|_{-1}^{1} - \int_{-1}^{1} f(x)(-2xg(x) + (1-x^2)g'(x))dx \\
+ \int_{-1}^{1} xf(x)g(x)dx \\
= \int_{-1}^{1} f(x) \left( (x^2-1)g'(x) + 3xg(x) \right) dx
\]

Thus, \( A^*g(x) = (x^2-1)g'(x) + 3xg(x) \).

In the above examples the needed differential equation for the density \( p \) of \( \mu \) could be established quite easily. One might ask if there is a universal way of deriving such an equation. At least, if one knows some homogeneous linear differential equation for the characteristic function \( \chi = \chi_\mu \) of \( \mu \) one can use Fourier inversion to obtain a linear differential equation for \( p \). This technique was applied in [GT05] to the density of the expected spectral distribution of a GUE random matrix (see Section 2.7 for an explanation of these notions). To be more concrete, denote by \( \mathcal{F} : L^1(\mathbb{R}, \mathcal{B}, \lambda) \rightarrow L^1(\mathbb{R}, \mathcal{B}, \lambda), \mathcal{F}f(t) := \int_{\mathbb{R}} e^{itx} f(x)dx \) the (probabilistic) Fourier transformation. Then

\[
\chi_\mu(t) = \int_{\mathbb{R}} e^{itx} \mu(dx) = \int_{\mathbb{R}} e^{itx} p(x)dx = \mathcal{F}p(t),
\]

and we have the following rules:

1. \( \frac{d^n}{dt^n} \mathcal{F}f(t) = i^n \mathcal{F}(x^n f)(t) \)
2. \( \mathcal{F} \left( \frac{d^n}{dx^n} f \right)(t) = (-i)^n t^n \mathcal{F}f(t) \)
2. Stein’s method for absolutely continuous univariate distributions

Since \( F \) is linear and one-to-one, these rules together with the linear differential equation for \( \chi \), imply the desired differential equation for the density \( p \).

Let us illustrate this approach by the following examples:

**Example 2.1.2. (a) The standard normal distribution \( N(0,1) \)**

It is well-known that in this case \( \chi(t) = e^{-t^2/2} \) and thus \( \chi'(t) + t\chi(t) = 0 \). Since \( \chi = Fp \) we conclude by the rules (1) and (2) that

\[
\chi'(t) = \frac{d}{dt} Fp(t) = if(xp)(t) \quad \text{and} \quad t\chi(t) = t(Fp)(t) = if(p')(t),
\]

so that by linearity and injectivity of \( F \) we have

\[ 0 = if(xp + p') \quad \text{and} \quad 0 = p'(x) + xp(x). \]

From that we can proceed as above to obtain the well-known Stein characterization for \( N(0,1) \).

**Example 2.1.2. (b) The Beta distributions on \([-1,1]\)**

This class of distributions is defined via the densities \( p_{\alpha,\beta}(x) = C(\alpha,\beta)(1-x)^{\alpha-1}(1+x)^{\beta}(1,1)(x) \) for \( \alpha,\beta > -1 \), where \( C(\alpha,\beta) = \frac{\Gamma(\alpha+\beta+2)}{2^\alpha\Gamma(\alpha+1)\Gamma(\beta+1)} \) (see Section 2.5). In the paper [Led04] the author finds a second order linear differential equation for the moment generating function of the distribution \( \mu_{\alpha,\beta} \) corresponding to \( p_{\alpha,\beta} \). This immediately yields the following differential equation for its characteristic function \( \chi_{\alpha,\beta} = Fp_{\alpha,\beta} \):

\[
t\chi''_{\alpha,\beta}(t) + (\alpha + \beta + 2)\chi'_{\alpha,\beta}(t) + (t + i(\alpha - \beta)) \chi_{\alpha,\beta}(t) = 0 \quad (2.1)
\]

By the rules (1) and (2) we obtain

\[
t\chi''_{\alpha,\beta}(t) = (-i)F \left( \frac{d}{dx} (x^2 p_{\alpha,\beta}) \right) (t) = -iF \left( 2xp_{\alpha,\beta} + x^2 p'_{\alpha,\beta} \right) (t),
\]

\[
\chi'_{\alpha,\beta}(t) = iF (xp_{\alpha,\beta}) (t) \quad \text{and} \quad t\chi_{\alpha,\beta}(t) = iF (p'_{\alpha,\beta}) (t)
\]

Thus, from (2.1) we conclude that

\[
0 = -iF \left( 2xp_{\alpha,\beta} + x^2 p'_{\alpha,\beta} \right) + (\alpha + \beta + 2)iF (xp_{\alpha,\beta}) + iF (p'_{\alpha,\beta}) + i(\alpha - \beta)F (p_{\alpha,\beta})
\]

\[
= -iF \left( 2xp_{\alpha,\beta} + x^2 p'_{\alpha,\beta} - (\alpha + \beta + 2)xp_{\alpha,\beta} - p'_{\alpha,\beta} + (\beta - \alpha)p_{\alpha,\beta} \right).
\]
2.1. Principles of finding a Stein type characterization

Since $\mathcal{F}$ is one-to-one, we thus have

$$0 = Ap_{\alpha,\beta} := (x^2 - 1)p'_{\alpha,\beta}(x) - (\alpha + \beta)xp_{\alpha,\beta}(x) + (\beta - \alpha)p_{\alpha,\beta}(x).$$

Integration by parts yields in the usual way the adjoint $A^*$ of $A$, which then serves as a Stein operator for $\mu_{\alpha,\beta}$ and which in this case is given by

$$A^*g(x) = (1 - x^2)g'(x) - (\alpha + \beta + 2)yg(x) + (\beta - \alpha)g(x).$$

This is the starting point for Stein’s method for Beta distributions, which is developed in Section 2.5. Note that the semi-circle distribution from Example 2.1.1 (b) is the special case $\alpha = \beta = 1/2$ of the Beta family.

2.1.2. The density approach

The differential equation approach, although being universal, has the drawback that one in the first place has to know a differential equation for the density or the characteristic function of the distribution $\mu$. In critical cases the so-called density approach might be more easily applied. This approach was first explored in [SDHR04] in the context of exchangeable pairs. The method of exchangeable pairs within this approach was then considerably extended by proposing a convenient regression property (see Subsection 2.3.1) independently in [EL10] and [CS11] and was applied to prove non-central limit theorems for the total magnetization in the critical temperature regime of a Curie-Weiss model.

In a nutshell, the density approach goes as follows: Suppose that $\text{supp}(\mu) = (a, b)$ and that the density $p$ is positive and absolutely continuous on every compact interval $[c, d] \subseteq (a, b)$. Let $\psi(x) := \frac{d}{dx} \log p(x) = \frac{p'(x)}{p(x)}$.

Then, under some regularity conditions on $p$ (see [SDHR04] for details), a real-valued random variable $X$ has distribution $\mu$ if and only if for all functions $f$ in an explicitly given class $\mathcal{F}$ (which depends on $\mu$) we have

$$E[f'(X) + \psi(X)f(X)] = f(b-)p(b-) - f(a+)p(a+),$$

where $p(b-), f(b-)$ and $p(a+), f(a+)$ denote the right and left limits of $p, f$, respectively, which are assured to exist by the regularity conditions on $p$ and the definition of the class $\mathcal{F}$ in [SDHR04]. This Stein characterization is proved in [SDHR04] and the proof reveals, that it suffices to consider those functions $f \in \mathcal{F}$ such that $\lim_{x \searrow a} f(x) = 0 = \lim_{x \nearrow b} f(x)$. Denoting this subclass by $\mathcal{F}'$, we have the following alternative characterization for $\mu$:
2. Stein’s method for absolutely continuous univariate distributions

**Theorem 2.1.3.** A real-valued random variable $X$ has distribution $\mu$ if and only if for all $f \in \mathcal{F}'$ we have

$$E[f'(X) + \psi(X)f(X)] = 0$$  \hfill (2.2)

With the Stein characterization and, hence, the Stein identity at hand, for a given Borel-measurable test function $h : \mathbb{R} \to \mathbb{R}$ with $E|h(Z)| < \infty$ we have the corresponding Stein equation

$$f'(x) + \psi(x)f(x) = h(x) - \mu(h),$$  \hfill (2.3)

to be solved for $f$ and where one usually only seeks a function $f$ of $x \in (a,b)$. Here we have written $\mu(h) = \int_{\mathbb{R}} h(x)d\mu(x) = E[h(Z)]$. Note that for general Borel-measurable $h$ it cannot be expected that there exists a solution $f$ which is differentiable on all of $(a,b)$ and satisfies (2.3) pointwise. Thus, we will understand a solution to be an almost everywhere differentiable and Borel-measurable function which satisfies (2.3) at all points $x \in (a,b)$ where it is in fact differentiable and contrary to the usual convention at the remaining points $x \in [a,b)$ we define $f'(x) := -\psi(x)f(x) + h(x) - \mu(h)$. This gives us a Borel-measurable function $f'$ on $(a,b)$ such that (2.3) holds for each $x \in (a,b)$. It is easily checked that a solution to (2.3) is given by

$$f_h(x) := \frac{1}{p(x)} \int_a^x (h(y) - \mu(h))p(y)dy = -\int_a^b (h(y) - \mu(h))p(y)dy$$  \hfill (2.4)

for $x \in (a,b)$. In most cases, this function $f_h$ may be extended continuously to $a$ and $b$. Typically, the solution $f_h$ has nice boundedness and smoothness properties and since every other solution $f$ of (2.3) must have the form $f(x) = f_h(x) + \frac{c}{p(x)}$ (note that the solutions of the corresponding homogeneous equation are parameterized by $\frac{c}{p(x)}$, $c \in \mathbb{R}$) and the function $\frac{1}{p(x)}$ is often unbounded at the edge of $(a,b)$. Then, it follows that $f_h$ is the only bounded solution of (2.3).

In general, it is not to be expected that the left hand side of equation (2.3) makes sense for $x \in \mathbb{R} \setminus (a,b)$, because the density $p$ is zero outside of $(a,b)$ and hence the definition $\psi(x) := \frac{\psi(x)}{p(x)}$ is impossible. However, in some cases for $x \in (a,b)$ the function $\psi$ is given by some “analytical expression”, which is also valid for $x \in \mathbb{R} \setminus (a,b)$ and this may serve as a convenient definition for $\psi(x)$. This topic will be addressed further in 2.2.

Let us have a look at some examples.

**Example 2.1.4.** (a) The standard normal distribution $N(0,1)$

Since $p'(x) = -xp(x)$ we have $\psi(x) = -x$ and hence we have for all $f \in \mathcal{F}'$: ...
2.1. Principles of finding a Stein type characterization

\[ E[f'(Z) - Zf(Z)] = 0, \]

which is again the well-known Stein identity for \( N(0,1) \).

(b) The semi-circle distribution on \([-1,1]\)

Here, according to Example 2.1.1 (b), \((1-x^2)p'(x) = -xp(x)\) and hence \(\psi(x) = \frac{-x}{1-x^2}\) which is unbounded at the edge of \((-1,1)\). Hence, we obtain for all \(f \in \mathcal{F}'\):

\[ E\left[f'(Z) - \frac{Z}{1-Z^2} f(Z)\right] = 0 \]

(c) The exponential distribution \(\text{Exp}(\lambda), \lambda > 0\)

Here, \(p(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x)\) and thus \(\psi(x) = -\lambda\) for all \(x \in (0,\infty)\). So, for all \(f \in \mathcal{F}'\):

\[ E[f'(Z) - \lambda f(Z)] = 0 \]

Note, that the Stein identity in Example 2.1.4 (b) is different from the Stein identity \(E[(1-Z^2)p'(Z) - 3Zg(Z)] = 0\) which comes from the operator \(-A^*\) from Example 2.1.1 (b). A Stein identity or characterization can never be expected to be unique, but one might choose from several possible characterizations by the peculiarities of the desired application.

It should be noted that a precise theorem like Theorem 2.1.3 with an explicitly given class \(\mathcal{F}\) or \(\mathcal{F}'\) is in general not necessary to apply Stein’s method for distributional approximations. Indeed, suppose we would like to bound the distance of the distribution of a given random variable \(W\) to the distribution \(\mu\) of \(Z\) in terms of a class \(\mathcal{H}\) of test functions \(h\) such that both \(E[h(W)]\) and \(E[h(Z)]\) exist. That is, we aim at bounding \(d_{\mathcal{H}}(W,Z) := \sup_{h \in \mathcal{H}} |E[h(W)] - E[h(Z)]|\).

Then it is sufficient that for each \(h \in \mathcal{H}\) the solution \(f_h\) given by (2.4) is integrable with respect to the distribution of \(W\) and, of course, that \(|E[f'_h(W)] - \psi(W)f_h(W)|\) can be appropriately bounded. It is thus in general not necessary to exactly define the class \(\mathcal{F}\). The characterization merely serves as a heuristic hint that Stein’s technique of using the Stein equation and its solution might be fruitful.

Probably the most general bounds on the solutions \(f_h\) from (2.4) were given in [CS11] (see also [CGS11], Chapter 13, for an elaboration of the theory from the paper [CS11] but also [EL10] for densities \(p(x) \propto \exp(-a_k|x|^{2k})\), where \(a_k > 0\) is some given constant).
2. Stein’s method for absolutely continuous univariate distributions

2.1.3. The parametric approach by Ley and Swan

In the recent preprint [LS11b] Ley and Swan propose a universal approach of finding a Stein characterization for a given quite arbitrary distribution that depends on a (possibly multidimensional) parameter $\theta$ of interest. Their technique relies on differentiating a certain integral identity with respect to the parameter $\theta$ in the sense of distributions. Arguing that nearly every common discrete or absolutely continuous distribution involves such a parameter, for example as a location or scale parameter, they are not only able to re-obtain most of the formerly known Stein characterizations but also derive several new ones.

2.1.4. The Malliavin calculus approach

When applying Stein’s method for normal approximation to a given random variable $W$ it is often useful to rewrite the expected value $E[Wf(W)]$ from the identity (1.2) into a form, which contains the derivative $f'$ of $f$ rather than the function $f$ itself. This can be accomplished through certain coupling constructions, as for example the exchangeable pairs coupling or even more obviously by the zero-bias coupling. In the seminal paper [NP09] Nourdin and Peccati combine Stein’s method and Malliavin calculus to prove error bounds on the normal approximation of functionals of a given centered isonormal Gaussian process $X = \{X(h) : h \in h\}$, where $h$ is a real separable Hilbert space. More precisely, for $W = F(X)$ a smooth functional of the process $X$ they used operators from Malliavin calculus to rewrite

\[
E[Wf(W)] = E[f'(W) \langle DW, -DL^{-1}W \rangle_h] = E[f'(W) E[\langle DW, -DL^{-1}W \rangle_h | W]],
\]

where $D$ is the Malliavin derivative and $L^{-1}$ is the pseudo-inverse of the generator $L$ of the Ornstein-Uhlenbeck semigroup on $L^2(\Omega, \sigma(X), P)$. This, among other results, allows them to derive a quantitative version of the so-called fourth moment theorem, formerly proved by Nualart and Peccati [NP05], which states that for a sequence of random variables, living in a fixed Wiener chaos of order at least 2 corresponding to $X$, convergence in distribution to $N(0,1)$ is equivalent to convergence of just the second and fourth moments to 1 and 3, respectively. In the same paper, they also treat approximation of such functionals $W$ by a centered Gamma distribution, which is possible because the Stein equation for the centered Gamma distribution with one parameter $\nu > 0$ is given by

\[
2(x + \nu)f'(x) - xf(x) = h(x) - E[h(Z)],
\]
2.2. The problem of support

with $Z$ having this Gamma distribution, and, hence, one may use (2.5) again to deal with $E[Wf(W)]$. In the paper [NV09] Nourdin and Viens prove, that a centered $\sigma(X)$-measurable random variable $Z \in D^{1,2}$, the domain of $D$, with values in $(a,b)$ satisfies the Stein identity

$$E[g_*(Z)f'(Z) - Zf(Z)] = 0$$

for all sufficiently well-behaved functions $f$. Here, $g_*$ is defined by

$$g_*(x) := E[\langle DZ, -DL^{-1}Z \rangle | Z = x].$$

If the function $g_*$ may be computed explicitly, this immediately yields a suitable Stein equation for the distribution $\mathcal{L}(Z)$. This was used in [EV12] by Eden and Viquez to bound the distance between $\mathcal{L}(Z)$ and the distribution of a given random variable $W = F(X)$ as above. There, the authors even allow $W$ to be an element of the more general Wiener-Poisson space.

2.2. The problem of support

Within the above approaches we have always started from a Stein identity for $\mu$ of the form $E[Lf(Z)] = 0$ for $f$ belonging to a rather rich class $\mathcal{F}$ of functions $(a,b) \to \mathbb{R}$, where $L$ was some differential operator whose domain also typically consists of functions $f : (a,b) \to \mathbb{R}$. For a given test function, the Stein equation $Lf = h - \mu(h)$, to be solved for $f$, hence also in the first place only makes sense for functions $f : (a,b) \to \mathbb{R}$. Now, given a real valued random variable $W$, which is supposed to be approximately distributed as $Z$ we cannot in the first place take the solution $f_h$ corresponding to the test function $h$ and use the identity

$$E[h(W)] - E[h(Z)] = E[Lf_h(W)]$$

in order to obtain bounds for the left hand side, because generally the function $Lf_h$ is only defined on $(a,b)$ but unless $(a,b) = \mathbb{R}$ we are not assured that $W(\omega) \in (a,b)$ for $P$-almost all $\omega \in \Omega$. This problem is not only of theoretical interest because there are important examples of sequences of random variables, each of which has unbounded support, but which converges in distribution to a random variable with bounded support. For example take an $n \times n$ Wigner random matrix (see Section 2.7 for this notion) and consider its ordered, real eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$, each single eigenvalue occurring as often as its multiplicity indicates. Let $I$ be uniformly distributed on $\{1, \ldots, n\}$, independent of the given Wigner matrix. Then, in [GT06] the authors show that the variable $\lambda_I$ has the expected spectral distribution of the Wigner matrix and hence, by Wigner’s semi-circle
2. Stein’s method for absolutely continuous univariate distributions

law, converges in distribution to the semi-circle distribution, which, of course, has bounded support, as $n$ goes to infinity. But, depending on the distribution of the entries of the Wigner matrix, the variable $\lambda_I$ may of course have support $\mathbb{R}$ for each fixed $n$.

There are at least two suggestions of how to overcome this problem:

(1) One could replace $W$ by a random variable $\tilde{W}$ which has values in $(a,b)$ and which coincides with $W$ for all $\omega$ such that $W(\omega) \in (a,b)$ and use the triangle inequality:

$$|E[h(W)] - E[h(Z)]| \leq |E[h(W)] - E[h(\tilde{W})]| + |E[h(\tilde{W})] - E[h(Z)]|$$

The first term on the right hand side might be easily bounded since the probability distribution of $W$ will nearly be concentrated on $(a,b)$ if $W$ is approximately distributed as $Z$. For the second term the Stein solution $f_h$ on $(a,b)$ may be used and it can be written as $E[Lf_h(\tilde{W})]$. However, this approach has an essential drawback. Namely, in many cases where Stein’s method proposes to be useful, $W$ has a certain arithmetic structure, for example it might be a sum, $W = \sum_{j=1}^n X_j$, of independent random variables $X_1, \ldots, X_n$. This structure does not generally carry over to $\tilde{W}$ and hence the application of Stein’s method, usually by couplings, might be much harder, if not infeasible, than for the original random variable $W$.

(2) The above example illustrates that it would be desirable to have a Stein equation and a corresponding solution, which are defined on the whole of $\mathbb{R}$. This problem is hard to tackle in full generality because the coefficients of the differential operator $L$ need not have a natural extension to $\mathbb{R} \setminus (a,b)$. Further and even more disillusioning, the Stein equation is motivated by the Stein identity and if we extend $L$ to $\tilde{L}$ acting on functions $\mathbb{R} \to \mathbb{R}$, then every such extension will provide a new Stein identity for $\mu$ because $Z \in (a,b)$ almost surely. Thus, we cannot hope to solve the problem of naturally extending the Stein operator $L$ on functions $\mathbb{R} \to \mathbb{R}$ in full generality.

Although we cannot propose a general procedure to overcome the problem of support, in many practical cases there is an extensions $\tilde{L}$ of $L$ which at least seems to be natural to us. For example, in the density approach, the coefficient $\psi$ of the Stein equation might be given by a certain “analytical expression” for $x \in (a,b)$ which perfectly makes sense for $x \in \mathbb{R} \setminus (a,b)$, too. As an easy example, take the exponential distribution $\text{Exp}(\lambda)$, where $\lambda > 0$. Here, $(a,b) = [0, \infty)$, $p(x) = \lambda e^{-\lambda x}1_{(0,\infty)}(x)$ and thus $\psi(x) = -\lambda$ for all $x \in (0, \infty)$. So the Stein equation can be naturally extended on $\mathbb{R}$.
2.3. The exchangeable pairs approach for absolutely continuous univariate distributions

The same often happens with Stein identities found by the differential equation approach. Take, for example, the semi-circle distribution on \([-1, 1]\) from Example 2.1.1 (b). In that case the Stein equation was \((1-x^2)g'(x) - 3xg(x) = h(x) - \mu(h)\) and, again, the functions \(\eta(x) := 1 - x^2\) and \(\gamma(x) := -3x\) can be naturally extended on \(\mathbb{R}\) by just the same expressions. This obviously also applies to the other Beta distributions.

2.3. The exchangeable pairs approach for absolutely continuous univariate distributions

The method of exchangeable pairs, which was first presented in Stein’s monograph [Ste86], is a cornerstone of Stein’s method and is still the most frequently used coupling. This is due to the wide applicability of standard couplings like the Gibbs sampler or making one time step in a reversible Markov chain, which generally yield exchangeable pairs. By definition, an exchangeable pair is a pair \((W, W')\) of random variables, defined on a common probability space, such that their joint distribution is symmetric, i.e. such that \((W, W') \overset{D}{=} (W', W)\). In [Ste86], in order to show that a given real-valued random variable \(W\) is approximately standard normally distributed, Stein proposes the construction of another random variable \(W'\), a small random perturbation of \(W\), on the same space as \(W\) such that \((W, W')\) forms an exchangeable pair and additionally the following linear regression property holds:

\[
E[W' - W | W] = -\lambda W
\] (2.6)

Here, \(\lambda \in (0, 1)\) is a constant which is typically close to zero for conveniently chosen \(W'\). If this condition is satisfied, then the distributional distance of \(L(W)\) to \(N(0, 1)\) can be efficiently bounded in various metrics, including the Kolmogorov and Wasserstein metrics (see, e.g. [Ste86], [CS05] or [CGS11] for the common “plug-in theorems”).

The range of examples to which this technique could be applied was considerably extended in the work [RR97] of Rinott and Rotar who proved normal approximation theorems allowing the linear regression property to be satisfied only approximately. Specifically, they assumed the existence of a random quantity \(R\), which is dominated by \(\lambda W\) in size, such that

\[
E[W' - W | W] = -\lambda W + R.
\] (2.7)

Note, that necessarily \(R\) is \(\sigma(W)\)-measurable and that unlike condition (2.6), condition (2.7) is not a true condition on the pair \((W, W')\) since we can always
2. Stein’s method for absolutely continuous univariate distributions

define \( R := E[W' - W|W] + \lambda W \) for each given constant \( \lambda > 0 \). However, the “plug-in theorems” in [RR97], [SS06] or [CGS11] clarify that \( R \) has to be of smaller order than \( \lambda \) in order to yield useful bounds. Since \( W \) is supposed to have a “true” distributional limit and, hence, \( \lambda^{-1}E[|R|] \) should be \( o(1) \), it follows that both, \( \lambda \) and \( R \), are at least asymptotically unique (see also the introduction of [RR09] for the discussion of this topic).

2.3.1. Exchangeable pairs and the density approach

If the non-normal target distribution \( \mu \) satisfies the conditions of the density approach in 2.1.2 and if one intends to use an exchangeable pairs coupling in order to bound the distributional distance of a given random variable \( W \) to \( \mu \) it might not be clear from the outset, which condition to substitute for the linear regression property (2.6) or (2.7). It was a crucial observation, independently made by Eichelsbacher and Löwe in [EL10] and by Chatterjee and Shao in [CS11], that the appropriate condition is

\[
E[W' - W|W] = \lambda \psi(W) + R, \tag{2.8}
\]

where, again \( \lambda > 0 \) is constant and \( R \) is of smaller order than \( \lambda \psi(W) \). Although leading to the same theoretical conclusion and comparable “plug-in theorems”, the papers [EL10] and [CS11] address the problem from two different directions, which is quite interesting to notice: On the one hand, in [EL10] the authors start with the target distribution \( \mu \) and with a random variable \( W \) (in the applications the magnetization in a statistical mechanics model) which they suppose to be distributed nearly as \( \mu \) and arrive precisely at condition (2.8) with \( \lambda \approx E[(W' - W)^2]/2 \). On the other hand, the authors in [CS11] (see also the book [CGS11]) start with an exchangeable pair \((W, W')\) which satisfies the condition

\[
E[W' - W|W] = g(W) + R, \tag{2.9}
\]

where, again, the term \( R \) on the right hand side is negligible. Then, they proceed exactly the other way around and define a density \( p \) which has as logarithmic derivative the function \( c_0 \cdot g \), where \( c_0 \) is a constant which they also discuss and which should satisfy \( c_0 \approx 2/E[(W' - W)^2] \). Thus \( c_0 \approx 1/\lambda \).

In order to give a flavour, we present parts of Theorem 2.4 from [EL10].

**Theorem 2.3.1.** Let the density \( p \) be positive and absolutely continuous on \((a, b)\). Suppose that there exist positive real constants \( c_1, c_2 \) and \( c_3 \) such that for any Lipschitz-continuous function \( h : (a, b) \to \mathbb{R} \) with minimal Lipschitz constant \( \|h'\|_\infty \), the solution \( f_h \) given by (2.4) satisfies
2.4. Modification of Stein’s equation by exchangeable pairs

\[ \|f_n\|_\infty \leq c_1 \|h\|_\infty, \quad \|f'_n\|_\infty \leq c_2 \|h'\|_\infty \quad \text{and} \quad \|f''_n\|_\infty \leq c_3 \|h''\|_\infty, \]

where, for a function \( f : (a, b) \to \mathbb{R} \) we let \( \|f\|_\infty \) be its essential supremum norm.

Then, for any exchangeable pair \((W, W')\) satisfying condition (2.8) we have

\[ |E[h(W)] - \mu(h)| \leq \|h'\|_\infty \left( c_2 E\left[ 1 - \frac{1}{2\lambda} E\left[ (W - W')^2 |W| \right] \right] + \frac{c_3}{4\lambda} E\left[ |W' - W|^3 \right] + \frac{c_1}{\lambda} \sqrt{E[R^2]} \right). \]

Remark 2.3.2. (a) From the third term of the bound (2.10) we see that, in fact, \( R \) must be of smaller order than \( \lambda \) in order for the bound to be useful.

(b) From the first term on the right hand side of (2.10) we can conclude that \( \lambda \) should indeed be such that \( E[(W' - W)^2] \approx 2\lambda \), if the bound is supposed to be accurate.

(c) The first term appearing in the bound on the right hand side of (2.10) is usually interpreted such, that the random variable \( E[(W' - W)^2 |W|)/2\lambda \) must “obey a law of large numbers” to obtain decreasing bounds. Bounding this term is often decisive for the success of applying Theorem 2.3.1.

(d) The conclusion that is often drawn from the theory developed in [EL10] and [CS11] is, that the distribution of a random variable \( W \) is close to \( \mu \) if the regression property (2.8) is satisfied and the random quantity \( E[(W - W')^2 |W|)/2\lambda \) obeys a law of large numbers. The fact, that in some important examples, Condition (2.8) does not hold with a negligible remainder term \( R \) motivates our development of a different Stein identity than (2.2) and Stein equation than (2.3) for the distribution \( \mu \) which are sometimes more useful. This will be done in Section 2.4.

2.4. Modification of Stein’s equation by exchangeable pairs

Having discussed the method of exchangeable pairs within the density approach in 2.3.1, we now address the problem, that condition (2.8) with negligible remainder term \( R \) is in some examples not satisfied by an exchangeable pair, which, however, appears natural to us for our approximation problem. For example, in many situations where the exchangeable pair \((W, W')\) is constructed via the Gibbs sampler, we have a regression property of the form (see Proposition A.2.3 (b))
2. Stein’s method for absolutely continuous univariate distributions

\[ E[W' - W|W] = \lambda \left(-c(W - E[Z])\right) + R, \quad (2.11) \]

where \( \lambda, c > 0 \) are constants (the reason why \( \lambda \) and \( c \) are not subsumed into a single constant will become clear later on) and where, again, \( R \) is a negligible remainder. Here, again \( Z \sim \mu \). Following the paper [CS11], condition (2.11) suggests approximating \( W \) with a normal distribution with mean \( E[Z] \). But there are situations, where the exchangeable pair \((W, W')\) is good, meaning that the difference \(|W' - W|\) is ‘small’, condition (2.11) is satisfied and where we know that \( W \) is approximately distributed as a non-normal random variable \( Z \sim \mu \) and so the normal approximation is inappropriate (see, e.g. Proposition 2.6.1). In general, this means that the law of large numbers, discussed in Remark 2.3.2 (c) and (d) cannot hold. This observation turns out to be even more dramatic, since the first term on the right hand side of (2.10) does not depend on the target distribution (except for the constant \( c_2 \)). So we can conclude, that the density approach itself must fail for the exchangeable pair \((W, W')\), irrespective of the target distribution we have in mind.

These observations motivate a new version of Stein’s method, that allows for a more general regression property.

Suppose, that an appropriately chosen exchangeable pair \((W, W')\) satisfies the following general regression property:

\[ E[W' - W|W] = \lambda \gamma(W) + R, \quad (2.12) \]

where \( \lambda > 0 \) is constant, \( \gamma \) is a measurable function, which is defined on \( \mathbb{R} \) if we only know that \( W \) is real-valued and which may only be defined on \((a, b)\) if we know that \( P(W \in (a, b)) = 1 \) and where \( R \) is a negligible remainder term. We will see, that it will be advantageous if the term \( \gamma(x) \cdot g(x) \) appears in the “new” Stein equation, where \( g \) is the unknown solution of this equation. So we make the following ansatz for the Stein identity:

\[ E[\eta(Z)g'(Z) + \gamma(Z)g(Z)] = 0, \quad (2.13) \]

where \( \eta \) is another function with the same domain as \( \gamma \), which still has to be defined in such a way that 2.13 characterizes \( \mu \).

Starting from the Stein identity (2.13) our aim is to identify the function \( \eta \). If this approach is successful, the Stein equation corresponding to a measurable function \( h \) will be

\[ \eta(x)g'(x) + \gamma(x)g(x) = h(x) - \mu(h). \quad (2.14) \]
2.4. Modification of Stein’s equation by exchangeable pairs

Let $f_h$ be the solution (2.4) to the equation (2.3). For the solution $g_h$ of (2.14) we make the ansatz $g_h(x) = \alpha(x)f_h(x)$ for some (sufficiently smooth) function $\alpha$. We let $\tilde{h} = h - \mu(h)$ and obtain

$$
\eta(x)\gamma'_{\tilde{h}}(x) = \eta(x)(\alpha'(x)f_h(x) + \alpha(x)f'_{\tilde{h}}(x))
= \eta(x)\alpha'(x)f_h(x) + \eta(x)\alpha(x)(\tilde{h}(x) - \psi(x)f_h(x))
= \eta(x)\alpha'(x)f_h(x) + \eta(x)\alpha(x)\tilde{h}(x) - \eta(x)\psi(x)g_h(x)
= \tilde{h}(x) - \gamma(x)g_h(x).
$$

(2.15)

For this identity to hold, irrespective of the test function $h$, it must be the case that $\alpha(x) = \frac{1}{\eta(x)}$ (particularly $\eta$ must be differentiable at least almost everywhere) and hence

$$
\alpha'(x) = \frac{-\eta'(x)}{\eta(x)^2} = -\frac{\eta'(x)}{\eta(x)}\alpha(x).
$$

(2.16)

Plugging this into (2.15) we obtain

$$
\eta(x)\gamma'_{\tilde{h}}(x) = -\eta'(x)g_h(x) - \eta(x)\psi(x)g_h(x) + \tilde{h}(x).
$$

This equals $\tilde{h}(x) - \gamma(x)g_h(x)$ if and only if $\eta$ satisfies the ordinary differential equation

$$
\eta'(x) = \gamma(x) - \psi(x)\eta(x).
$$

(2.17)

This is a first order linear differential equation, which can of course be solved explicitly by the method of variation of the constant. Since, for $x \in (a, b)$, we have $\psi(x) = \frac{d}{dx}(\log p)(x)$ and thus, for any $d \in (a, b)$,

$$
\int_d^x \psi(t)dt = \log p(x) - \log p(d),
$$

the solution $\eta_0$ of the homogeneous equation corresponding to (2.17) with $\eta_0(d) = 1$ is given by

$$
\eta_0(x) = \exp \left( - \int_d^x \psi(t)dt \right) = \frac{p(d)}{p(x)}.
$$

Now, since $\eta(x) = C(x) \cdot \eta_0(x)$ we obtain that

$$
\eta'(x) = C'(x)\eta_0(x) - C(x)\psi(x)\eta_0(x) = C'(x)\eta_0(x) - \psi(x)\eta(x)
$$

which equals $\gamma(x) - \psi(x)\eta(x)$ if and only if $C'(x)\eta_0(x) = \gamma(x)$. Hence,
2. Stein’s method for absolutely continuous univariate distributions

\[ C(x) = C(d) + \int_d^x \gamma(t) \frac{1}{\eta_0(t)} dt = C(d) + \frac{1}{p(d)} \int_d^x \gamma(t)p(t) dt. \]

Thus, we have

\[ \eta(x) = \left( C(d) + \frac{1}{p(d)} \int_d^x \gamma(t)p(t) dt \right) \frac{p(d)}{p(x)} = \left( C(d) + \int_d^x \gamma(t)p(t) dt \right) \frac{1}{p(x)}. \]

First choosing \( C(d) = 0 \) and then letting \( d \downarrow a \) we obtain

\[ \eta(x) = \frac{1}{p(x)} \int_a^x \gamma(t)p(t) dt, \tag{2.18} \]

at least, if \( \int_a^b |\gamma(t)|p(t)dt = E[|\gamma(Z)|] < \infty \). But this is a very natural condition to hold, since the function \( \gamma \) was motivated by the regression property (2.12) and so, if the random variables \( W \) and \( W' \) are integrable we obtain that both sides of (2.12) must be \( P \)-integrable and, in fact

\[ E[\lambda \gamma(W) + R] = E[E[W' - W|W]] = E[W'] - E[W] = 0. \]

Neglecting the remainder term \( R \) we thus see that \( E[\gamma(W)] \) should exist and, in fact, be close to zero. So since \( W \overset{D}{=} Z \) we find it reasonable that \( E[\gamma(Z)] \) exists and even equals zero. Furthermore, it is a matter of routine to check, that \( \eta \) as given in (2.18) indeed still satisfies (2.17).

The above calculations starting with (2.13) were rather formal but crucial for the motivation and understanding of our approach.

In order to derive precise results, we will first formulate some technical conditions, additional to the general conditions in this section, on the target distribution \( \mu \), mainly through its Lebesgue density \( p \) and on the desired coefficient \( \gamma \) of the unknown function \( g \) in (2.14).

**Condition 2.4.1.** The density \( p \) is positive on the interval \( (a, b) \) and absolutely continuous on every compact interval \( [c, d] \subseteq (a, b) \).

**Condition 2.4.2.** The function \( \gamma : (a, b) \to \mathbb{R} \) is such that

(i) \( \gamma \) is continuous on \( (a, b) \)

(ii) \( \gamma \) is strictly decreasing on \( (a, b) \)

(iii) \( \int_a^b |\gamma(t)|p(t)dt < \infty \) and in fact \( E[\gamma(Z)] = \int_a^b \gamma(t)p(t)dt = 0 \)

(iv) There is a unique \( x_0 \in (a, b) \) with \( \gamma(x_0) = 0 \).
2.4. Modification of Stein’s equation by exchangeable pairs

Remark 2.4.3. Note that in Condition 2.4.2 (iv) is actually implied by (i), (ii) and (iii) and the intermediate value theorem. Furthermore, (iv) implies that $\gamma$ is positive on $(a, x_0)$ and is negative on $(x_0, b)$.

By item (iii) in Condition 2.4.2 we can define the function $I : (a, b) \to \mathbb{R}$ by $I(x) := \int_a^x \gamma(t)p(t)\,dt$ which is continuously differentiable on $(a, b)$ if $\gamma$ satisfies Condition 2.4.2.

Proposition 2.4.4. Under Conditions 2.4.1 and 2.4.2 the function $I$ has the following properties:

(a) $I(x) = -\int_x^b \gamma(t)p(t)\,dt$

(b) $I(x) > 0$ for each $x \in (a, b)$

(c) $I$ is strictly increasing on $(a, x_0)$ and strictly decreasing on $(x_0, b)$ and hence attains its global maximum at $x_0$.

Proof. Of course, (a) is immediately implied by item (iii) from Condition 2.4.2. To prove (b) and (c) first observe that by (iii) we have $\lim_{x \to a^+} I(x) = 0 = \lim_{x \to b^-} I(x)$. Furthermore, $I'(x) = \gamma(x)p(x)$ is positive on $(a, x_0)$ and negative on $(x_0, b)$ implying the results.

We begin developing Stein’s method for the distribution $\mu$ satisfying Condition 2.4.1 and the coefficient $\gamma$ of the unknown function $g$ in (2.14), which we assume to fulfill Condition 2.4.2. We define the function $\eta : (a, b) \to \mathbb{R}$ by (2.18) and for a given Borel-measurable test function $h$ with $E[|h(Z)|] < \infty$ we consider the Stein equation (2.14). In most cases of practical interest we will have $\lim_{x \to a^+} \eta(x) = 0 = \lim_{x \to b^-} I(x)$. Furthermore, $I'(x) = \gamma(x)p(x)$ is positive on $(a, x_0)$ and negative on $(x_0, b)$ implying the results.

Proposition 2.4.4. Under Conditions 2.4.1 and 2.4.2 the function $\eta$ has the following properties:

(a) $\eta$ is positive on $(a, b)$, absolutely continuous on every compact subinterval $[c, d] \subseteq (a, b)$ and $\eta'(x) = \gamma(x) - \psi(x)\eta(x)$ for $\lambda$-almost all $x \in (a, b)$.

(b) $\lim_{x \to a^+} \eta(x)p(x) = \lim_{x \to b^-} \eta(x)p(x) = 0$

(c) If $\lim_{x \to a^+} p(x) = 0$, then $\lim_{x \to a^+} \eta(x) = \frac{\gamma(a)}{\lim_{x \to a^+} \psi(x)}$, if this limit exists.

(d) If $\lim inf_{x \to a^+} p(x) \in (0, \infty) \cup \{\infty\}$ then $\lim_{x \to a^+} \eta(x) = 0$

31
2. Stein’s method for absolutely continuous univariate distributions

(e) If \( \lim_{x \uparrow b} p(x) = 0 \), then \( \lim_{x \uparrow b} \eta(x) = \frac{\gamma(b)}{\lim_{x \uparrow b} \psi(x)} \), if this limit exists.

(f) If \( \lim \inf_{x \uparrow b} p(x) \in (0, \infty) \cup \{\infty\} \) then \( \lim_{x \uparrow b} \eta(x) = 0 \).

Proof. The first part of (a) follows from the fact that \( I \) is positive on \((a,b)\) and \( C^1 \) on \((a,b)\) and hence absolutely continuous on \([c,d]\) and that \( p \) is also absolutely continuous and bounded below by a positive constant on \([c,d]\). The rest of (a) has already been observed. Items (b), (d) and (f) follow immediately from the properties of the function \( I \) in Proposition 2.4.4. To prove (c), we use de l’Hôpital’s rule (see Theorem A.1.5) to derive

\[
\lim_{x \downarrow a} \eta(x) = \lim_{x \downarrow a} \frac{I(x)}{p(x)} = \lim_{x \downarrow a} \frac{\gamma(x)p(x)}{p'(x)} = \frac{\gamma(a)}{\lim_{x \downarrow a} \psi(x)}.
\]

In a similar way one can prove (e). \( \Box \)

Remark 2.4.6. In most practical cases where \( a > -\infty \) and/or \( b < \infty \) we will not only have \( \lim_{x \downarrow a} \eta(x) = \lim_{x \uparrow b} \eta(x) = 0 \) if \( \lim_{x \downarrow a} p(x) \in (0, \infty) \cup \{\infty\} \) and \( \lim_{x \uparrow b} p(x) \in (0, \infty) \cup \{\infty\} \). The following “Mill’s ratio” condition on the density \( p \) and the corresponding distribution function \( F \) is often satisfied and will also yield this behaviour of \( \eta \) at the boundary of \( \text{supp}(\mu) \) if \( a > -\infty \) and/or \( b < \infty \).

Condition 2.4.7. The density \( p \) of \( \mu \) satisfies all the properties from Condition 2.4.1 and also the following:

(i) If \( a > -\infty \), then \( \lim_{x \uparrow a} \frac{F(x)}{p(x)} = 0 \).

(ii) If \( b < \infty \), then \( \lim_{x \uparrow b} \frac{1-F(x)}{p(x)} = 0 \).

Remark 2.4.8. (a) Condition 2.4.7 is always satisfied if the density \( p \) is bounded away from zero in suitable neighbourhoods of \( a \) and \( b \).

(b) Assume that both, \( a > -\infty \) and \( b < \infty \) and that \( \lim_{x \downarrow a} p(x) = \lim_{x \uparrow b} p(x) = 0 \). Then Condition 2.4.7 is satisfied, if there is a \( \delta > 0 \) such that \( p \) is increasing on \((a,a+\delta)\) and decreasing on \((b-\delta,b)\). This is easily seen by the inequality

\[
F(x) = \int_a^x p(t)dt \leq p(x)(x-a),
\]

valid for \( x \in (a,a+\delta) \) and a similar one for the right end point \( b \).

(c) Suppose that \( a > -\infty \) and \( b < \infty \), that \( \lim_{x \downarrow a} p(x) = \lim_{x \uparrow b} p(x) = 0 \) and that there is a \( \delta > 0 \) such that \( p \) is convex on \((a,a+\delta)\) and on \((b-\delta,b)\). Then the assumptions of (b) and hence Condition 2.4.7 is satisfied. In fact,
2.4. Modification of Stein's equation by exchangeable pairs

First we can extend \( p \) to a continuous and convex function on \([a,b]\) by setting \( p(a) := 0 \). Now, let \( a < x < y < a + \delta \). Then, there exists a \( \lambda \in (0,1) \) with \( x = \lambda a + (1-\lambda)y \) and by convexity we have:

\[
p(y) - p(x) = p(y) - p(\lambda a + (1-\lambda)y) \geq p(y) - \lambda p(a) - (1-\lambda)p(y) = \lambda p(y) > 0
\]

Thus, \( p \) is strictly increasing on \((a,a+\delta)\). Similarly, one shows, that \( p \) is strictly decreasing on \((b-\delta,b)\), if \( p \) is convex there.

(d) If \( p \) is analytic at \( a \) and \( b \), then Condition 2.4.7 is also satisfied. Indeed, if there is an \( r > 0 \) such that \( p(x) = \sum_{k=0}^{\infty} c_k (x-a)^k \) for all \( x \in (a, a+r) \), then the function \( f : (a-r,a+r) \to \mathbb{R} \) with \( f(x) := \sum_{k=0}^{\infty} c_k (x-a)^k \) is well-defined. Let \( n_0 := \min \{ k \geq 0 : c_k \neq 0 \} \). Then \( n_0 < \infty \) since \( \text{supp}(\mu) = (a,b) \).

If \( n_0 = 0 \) and hence \( f(a) = c_0 = \lim_{x \searrow a} p(x) \neq 0 \), then there is nothing to show. Otherwise, we have \( p(x) = (x-a)^{n_0} \sum_{k=n_0}^{\infty} c_k (x-a)^{k-n_0} \) and \( p'(x) = (x-a)^{n_0-1} \sum_{k=n_0}^{\infty} kc_k (x-a)^{k-n_0} \) for \( x \in (a,a+r) \) and hence, by de l'Hôpital's rule

\[
\lim_{x \searrow a} \frac{F(x)}{p(x)} = \lim_{x \searrow a} \frac{p(x)}{p'(x)} = \lim_{x \searrow a} \frac{\sum_{k=n_0}^{\infty} c_k (x-a)^{k-n_0}}{\sum_{k=n_0}^{\infty} kc_k (x-a)^{k-n_0}} = \frac{c_{n_0}}{n_0 c_{n_0}} \lim_{x \searrow a} (x-a) = 0.
\]

\( \Box \)

The following proposition provides the results which were announced in Remark 2.4.6.

**Proposition 2.4.9.** Assume Condition 2.4.7. Then the function \( \eta \) vanishes at the finite end points of the support \((a,b)\) of \( \mu \), i.e. if \( a > -\infty \), then \( \lim_{x \searrow a} \eta(x) = 0 \) and if \( b < \infty \), then \( \lim_{x \nearrow b} \eta(x) = 0 \). Hence, we may extend \( \eta \) to a continuous function on \((a,b)\) by letting \( \eta(a) := \eta(b) := 0 \).

**Proof.** Suppose, that \( a > -\infty \). Then, by the positivity of \( I \) and the monotonicity of \( \gamma \), for \( a < x < b \):

\[
0 < I(x) = \int_{a}^{x} \gamma(t)p(t)dt < \gamma(a) \int_{a}^{x} p(t)dt = \gamma(a)F(x)
\]

Hence,
2. Stein’s method for absolutely continuous univariate distributions

\[ 0 \leq \lim \inf_{x \rightarrow a} \eta(x) \leq \lim \sup_{x \rightarrow a} \eta(x) = \lim \sup_{x \rightarrow a} \frac{I(x)}{p(x)} \leq \gamma(a) \lim_{x \rightarrow a} \frac{F(x)}{p(x)} = 0, \]

so that \( \lim_{x \rightarrow a} \eta(x) = 0 \). Since \( \eta \) is continuous, it follows that \( \eta(a) = 0 \).

The proof of \( \eta(b) = 0 \) for finite \( b \) is similar by using the representation \( I(x) = -\int_{x}^{b} \gamma(t)p(t)dt \) and is therefore omitted.

From our above heuristic calculations we know that the function \( g_{h} : (a, b) \rightarrow \mathbb{R} \) with

\[ g_{h}(x) := \frac{f_{h}(x)}{\eta(x)} = \frac{1}{p(x) \eta(x)} \int_{a}^{x} (h(t) - \mu(h))p(t)dt \]

\[ = -\frac{1}{p(x) \eta(x)} \int_{a}^{b} (h(t) - \mu(h))p(t)dt \quad (2.19) \]

solves the Stein equation (2.14) for \( x \in (a, b) \). This can also be proved directly by differentiation and the formula for \( g_{h} \) could also be derived by the method of variation of the constant, very similar to the derivation of the formula for \( \eta \). One merely has to know, that \( \log(p \cdot \eta) \) is a primitive function of \( \frac{\gamma}{\eta} \), which follows from (2.17) by a simple calculation. If we can show that \( g_{h} \) is bounded, then it will immediately follow that \( g_{h} \) is the only bounded solution of (2.14) since the solutions of the corresponding homogeneous equation are constant multiples of \( \frac{1}{p \eta} \) which is unbounded by Proposition 2.4.5 (a).

As is typical for Stein’s method, its success within the applications considerably depends on good bounds on the solutions \( g_{h} \) and their derivative(s), generally uniformly over some prominent class \( \mathcal{H} \) of test functions \( h \).

The next step will be to prove such bounds. It has to be mentioned that we cannot expect to derive concrete good bounds in full generality, but that sometimes further conditions have to be imposed either on the distribution \( \mu \) (e.g., through the density \( p \)) or on the coefficient \( \gamma \). Nevertheless, we will derive bounds involving functional expressions which can a posteriori be simplified, computed or further bounded for concrete distributions. So our abstract viewpoint will pay off. Moreover, some of our bounds will actually hold completely generally.

The next proposition contains bounds on the solutions \( g_{h} \) for bounded and Borel-measurable test functions \( h \).

**Proposition 2.4.10.** Assume Conditions 2.4.1 and 2.4.2 and let \( m \) be a median for \( \mu \). Then, for \( h : (a, b) \rightarrow \mathbb{R} \) Borel-measurable and bounded we have
2.4. Modification of Stein’s equation by exchangeable pairs

\[ \|g_h\|_\infty \leq \frac{\|h - \mu(h)\|_\infty}{2I(m)} = \frac{\|h - \mu(h)\|_\infty}{2 \int_a^m \gamma(t) p(t) dt}. \tag{2.20} \]

**Proof.** With \( \tilde{h} = h - \mu(h) \), since \( I = \eta \cdot p \), we have for \( a < x < b \)

\[ |g_h(x)| = \frac{|\int_a^x \tilde{h}(t) p(t) dt|}{|p(x)\eta(x)|} = \frac{|\int_a^x \tilde{h}(t) p(t) dt|}{I(x)} \leq \|\tilde{h}\|_\infty \frac{F(x)}{I(x)}. \]

Let \( M: (a, b) \to \mathbb{R} \) be given by \( M(x) := \frac{F(x)}{I(x)} \). By l’Hôpital’s rule (see Theorem A.1.5) we have

\[ \lim_{x \downarrow a} M(x) = \lim_{x \downarrow a} \frac{p(x)}{\gamma(x)p(x)} = \lim_{x \downarrow a} \frac{1}{\gamma(x)} = \frac{1}{\lim_{x \downarrow a} \gamma(x)} \]

which exists in \([0, \infty)\) by Condition 2.4.2. Here, we used the convention \( \frac{1}{\infty} = 0 \). Moreover,

\[ \lim_{x \uparrow b} M(x) = \frac{1}{\lim_{x \uparrow b} I(x)} = +\infty \]

again by Condition 2.4.2 and by Proposition 2.4.4. Furthermore, we have

\[ M'(x) = \frac{p(x)I(x) - p(x)\gamma(x)F(x)}{I(x)^2} = \frac{p(x)}{I(x)^2} \left( I(x) - \gamma(x)F(x) \right) > 0 \]

for each \( x \in (a, b) \) since by the positivity of \( p \) and because \( \gamma \) is strictly decreasing

\[ I(x) = \int_a^x \gamma(t)p(t) dt > \gamma(x) \int_a^x p(t) dt = \gamma(x)F(x). \]

Hence, \( M \) is strictly increasing and thus for each \( x \in (a, m) \):

\[ |g_h(x)| \leq \|\tilde{h}\|_\infty \frac{F(m)}{I(m)} = \frac{\|h - \mu(h)\|_\infty}{2I(m)}. \]

The same bound can be proved for \( x \in (m, b) \) by using the representation

\[ g_h(x) = -\frac{1}{I(x)} \int_x^b (h(t) - \mu(h))p(t) dt \]

and the fact that also \( 1 - F(m) = \frac{1}{2} \). \qed
2. Stein’s method for absolutely continuous univariate distributions

The following corollary specializes this result to the case that \( \gamma(x) = -c(x - E[Z]) \) and that \( \mu \) is symmetric with respect to its median, which is then equal to its expected value, \( m = E[Z] \), that is \( Z - m \overset{D}{=} m - Z \).

**Corollary 2.4.11.** In addition to Conditions 2.4.1 and 2.4.2 assume that the distribution \( \mu \) is symmetric with respect to \( m = E[Z] \) and that \( \gamma(x) = -c(x - E[Z]) \) for some positive constant \( c \). Then for each bounded and Borel-measurable test function \( h : (a,b) \to \mathbb{R} \) we have

\[
\|g_h\|_\infty \leq \frac{\|h - \mu(h)\|_\infty}{cE[|Z - m|]} \quad (2.21)
\]

**Proof.** In this case we clearly have \( I(m) = \frac{c}{2}E[|Z - m|] \) which implies the result by Proposition 2.4.10. \( \square \)

In the case that \( \mu = N(0,1) \) and \( c = 1 \) this result specializes to the well known bound \( \sqrt{\frac{2}{\pi}}\|h - \mu(h)\|_\infty \) (see [CGS11] or [CS05], e.g.).

**Remark 2.4.12.** In the formulation of Proposition 2.4.10 it might surprise that there is no bound mentioned for \( \|g'_h\|_\infty \). This is because, in general a bound of the form \( \|g'_h\|_\infty \leq C\|h\|_\infty \) does not exist with a finite constant \( C \). This will become clear in the case of the Beta distributions in Section 2.5.

Next, we will turn to Lipschitz continuous test functions \( h \). In contrast to bounded measurable test functions, there we will also be able to prove useful bounds for \( \|g'_h\|_\infty \). We will occasionally need the following two lemmas.

**Lemma 2.4.13.** Let \(-\infty \leq a < b \leq \infty \) and let \( \mu \) be a probability measure (not necessarily absolutely continuous with respect to \( \lambda \)) with \( \text{supp}(\mu) \subseteq [a,b] \). Let \( F \) be the distribution function corresponding to \( \mu \) and suppose that \( \int_a^b |x|d\mu(x) < \infty \). Then, for each \( x \in (a,b) \) we have

(a) \( \int_a^x F(t)dt = xF(x) - \int_{(a,x]} s d\mu(s) \)

(b) \( \int_x^b (1 - F(t))dt = \int_{(x,\infty)} sd\mu(s) - x(1 - F(x)) \)

**Proof.** By Fubini’s theorem we have
2.4. Modification of Stein’s equation by exchangeable pairs

\[ \int_a^x F(t)dt = \int_x^{-\infty} F(t)dt = \int_{-\infty}^{x} F(t)dt = \int_{-\infty}^{x} \left( \int_{-\infty}^{t} d\mu(s) \right)dt \]

\[ = \int_{-\infty}^{x} \left( \int_{s}^{x} dt \right) d\mu(s) = \int_{-\infty}^{x} (x-s)d\mu(s) \]

\[ = xF(x) - \int_{-\infty}^{x} s d\mu(s) \]

This proves (a). Similarly, we have

\[ \int_{x}^{b} (1-F(t))dt = \int_{x}^{\infty} (1-F(t))dt = \int_{(x,\infty)} \mu((t,\infty))dt \]

\[ = \int_{(x,\infty)} \int_{(t,\infty)} d\mu(s)dt = \int_{(x,\infty)} \int_{(x,s)} dt d\mu(s) \]

\[ = \int_{(x,\infty)} (s-x)d\mu(s) \]

\[ = \int_{(x,\infty)} s d\mu(s) - x(1-F(x)), \]

proving (b).

\[ \square \]

**Lemma 2.4.14.** Let \(-\infty \leq a < b \leq \infty\) and let \(\mu\) be a probability measure (not necessarily absolutely continuous with respect to \(\lambda\)) with \(\text{supp}(\mu) \subseteq (a,b)\). Let \(F\) be the distribution function corresponding to \(\mu\), let \(Z \sim \mu\) and let \(h : (a,b) \to \mathbb{R}\) be Lipschitz continuous with \(E[|h(Z)|] < \infty\). Then the following assertions hold true:

(a) For each \(y \in \mathbb{R}\) we have

\[ h(y) - \mu(h) = \int_{-\infty}^{y} F(s)h'(s)ds - \int_{y}^{\infty} (1-F(s))h'(s)ds. \]

(b) For each \(x \in (a,b)\) we have

\[ \int_{(a,x)} (h(y) - \mu(h))d\mu(y) = -\left(1-F(x)\right) \int_{a}^{x} F(s)h'(s)ds - F(x) \int_{x}^{b} (1-F(s))h'(s)ds. \]

37
2. Stein’s method for absolutely continuous univariate distributions

Proof. Since \( \mu \) is a probability measure we have by the fundamental theorem of calculus for Lebesgue integration and by Fubini’s theorem

\[
\begin{align*}
    h(y) - \mu(h) &= \int_{\mathbb{R}} \left( h(y) - h(t) \right) d\mu(t) = \int_{\mathbb{R}} \left( \int_t^y h'(s) ds \right) d\mu(t) \\
    &= \int_{(-\infty,y]} \left( \int_t^y h'(s) ds \right) d\mu(t) - \int_{(y,\infty]} \left( \int_t^y h'(s) ds \right) d\mu(t) \\
    &= \int_{(-\infty,y]} \left( \int_{(-\infty,s]} d\mu(t) \right) h'(s) ds - \int_{(y,\infty)} \left( \int_{(s,\infty]} d\mu(t) \right) h'(s) ds \\
    &= \int_{(-\infty,y]} F(s) h'(s) ds - \int_{(y,\infty)} (1 - F(s)) h'(s) ds \\
    &= \int_{-\infty}^y F(s) h'(s) ds - \int_{y}^\infty (1 - F(s)) h'(s) ds.
\end{align*}
\]

This proves (a). As to (b), we have using (a) and its proof

\[
\begin{align*}
    \int_{(a,x]} (h(y) - \mu(h)) d\mu(y) &= \int_{(-\infty,x]} (h(y) - \mu(h)) d\mu(y) \\
    &= \int_{(-\infty,x]} \left( \int_{(-\infty,y]} F(s) h'(s) ds - \int_{(y,\infty]} (1 - F(s)) h'(s) ds \right) d\mu(y) \\
    &= \int_{(-\infty,x]} \int_{(-\infty,y]} F(s) h'(s) ds d\mu(y) - \int_{(-\infty,x]} \int_{(y,\infty)} (1 - F(s)) h'(s) ds d\mu(y) \\
    &= \int_{(-\infty,x]} F(s) h'(s) \left( \int_{(s,\infty]} d\mu(y) \right) ds - \int_{(-\infty,x]} (1 - F(s)) h'(s) \left( \int_{(-\infty,s]} d\mu(y) \right) ds \\
    &= \int_{(-\infty,x]} F(s) h'(s)(F(x) - F(s)) ds - \int_{(-\infty,x]} F(s)(1 - F(s)) h'(s) ds \\
    - \int_{(x,\infty)} F(x)(1 - F(s)) h'(s) ds.
\end{align*}
\]
2.4. Modification of Stein’s equation by exchangeable pairs

\begin{align*}
&= \int_{-\infty}^{x} \left( F(s)F(x)h'(s) - F(s)^2h'(s) + F(s)^2h'(s) - F(s)h'(s) \right) ds \\
&\quad - F(x) \int_{x}^{\infty} (1 - F(s))h'(s) ds \\
&= -(1 - F(x)) \int_{-\infty}^{x} F(s)h'(s) ds - F(x) \int_{x}^{\infty} (1 - F(s))h'(s) ds \\
&= -(1 - F(x)) \int_{a}^{x} F(s)h'(s) ds - F(x) \int_{x}^{b} (1 - F(s))h'(s) ds,
\end{align*}

as claimed.

In order to apply Lemma 2.4.13 we need a further condition on the distribution \( \mu \) which guarantees that its expected value exists.

**Condition 2.4.15.** The density \( p \) is positive on the interval \( (a, b) \) and absolutely continuous on every compact interval \( [c, d] \subseteq (a, b) \). Furthermore, \( E[|Z|] = \int_{[a,b]} |x| p(x) dx < \infty \).

The following proposition includes bounds for both, \( g_h \) and \( g'_h \), when \( h \) is Lipschitz.

**Proposition 2.4.16.** Under Conditions 2.4.15 and 2.4.2 we have for any Lipschitz-continuous test function \( h : (a, b) \to \mathbb{R} \) and any \( x \in (a, b) \):

(a) \( |g_h(x)| \leq \|h'\|_{\infty} \int_{a}^{x} F(s)E[Z] - \int_{a}^{s} F(y)p(y)dy \)

(b) \( |g'_h(x)| \leq \|h'\|_{\infty} \int_{a}^{x} F(s)ds(\gamma(x) + \int_{a}^{s} (1 - F(s))ds)H(x) \)

Here, for \( x \in (a, b) \) the positive function \( H \) is defined by

\[ H(x) := I(x) - \gamma(x)F(x) = p(x)\eta(x) - \gamma(x)F(x). \]

**Proof.** First, we prove (a). Recall the representation

\[ g_h(x) = \frac{1}{I(x)} \int_{a}^{x} (h(y) - \mu(h))p(y)dy = \frac{1}{I(x)} \int_{(a,x]} (h(y) - \mu(h))d\mu(y). \]

By Lemmas 2.4.14 and 2.4.13 we thus obtain that
2. Stein’s method for absolutely continuous univariate distributions

\[ |I(x)g_h(x)| = \left| -(1 - F(x)) \int_a^x F(s)h'(s)ds - F(x) \int_x^b (1 - F(s))h'(s)ds \right| \leq \|h'\|_\infty \left( (1 - F(x)) \int_a^x F(s)ds + F(x) \int_x^b (1 - F(s))ds \right) \]

\[ = \|h'\|_\infty \left( (1 - F(x)) \left( xF(x) - \int_a^x sp(s)ds \right) + F(x) \left( -x(1 - F(x)) + \int_x^b sp(s)ds \right) \right) \]

\[ = \|h'\|_\infty \left( - \int_a^x sp(s)ds + F(x) \left( \int_a^x sp(s)ds + \int_x^b sp(s)ds \right) \right) \]

\[ = \|h'\|_\infty \left( F(x)E[Z] - \int_a^x yp(y)dy \right) , \]

implying (a).

Now, we turn to the proof of (b). By Stein’s equation (2.14) we obtain for \( x \in (a, b) \)

\[ g'_h(x) = \frac{1}{\eta(x)} \left( \tilde{h}(x) - \gamma(x)g_h(x) \right) , \quad (2.22) \]

where we have again written \( \tilde{h} = h - \mu(h) \). Using Lemma 2.4.14 again, we obtain

\[ g_h(x) = \frac{1}{\eta(x)} \left( \int_a^x F(s)h'(s)ds \left( 1 + \frac{\gamma(x)(1 - F(x))}{\eta(x)p(x)} \right) + \int_x^b (1 - F(s))h'(s)ds \left( -1 + \frac{\gamma(x)F(x)}{\eta(x)p(x)} \right) \right) \]

\[ = \int_a^x F(s)h'(s)ds \left( \frac{\eta(x)p(x) + \gamma(x)(1 - F(x))}{\eta(x)^2 p(x)} \right) \]

\[ + \int_x^b (1 - F(s))h'(s)ds \left( \frac{-\eta(x)p(x) + \gamma(x)F(x)}{\eta(x)^2 p(x)} \right) \quad (2.23) \]

We define the functions \( H, G : (a, b) \to \mathbb{R} \) by \( H(x) := I(x) - \gamma(x)F(x) = \eta(x)p(x) - \gamma(x)F(x) \) and \( G(x) := H(x) + \gamma(x) = I(x) + \gamma(x)(1 - F(x)) \). It was already observed in the proof of Proposition 2.4.10 that \( H \) is positive on \( (a, b) \).
Similarly we prove the positivity of $G$ on $(a,b)$: For $x$ in $(a,b)$ we have, since $p$ is positive and $\gamma$ is strictly decreasing

$$G(x) = I(x) + \gamma(x)(1-F(x)) = -\int_x^b \gamma(t)p(t)dt + \gamma(x)(1-F(x))$$

$$> -\gamma(x)(1-F(x)) + \gamma(x)(1-F(x)) = 0.$$  

By (2.23) we can thus bound

$$|g'(x)| \leq \|h''\|_{\infty} \left( \int_a^x F(s)ds \frac{G(x)}{\eta(x)^2p(x)} + \int_x^b (1-F(s))ds \frac{H(x)}{\eta(x)^2p(x)} \right),$$

which reduces to the bound asserted in (b).

**Remark 2.4.17.** In general, the term $S(x) := F(x)E[Z] - \int^x F(y)p(y)dy$ cannot be bounded uniformly in $x \in (a,b)$ unless $|\gamma|$ grows at least linearly in $x$.

For further reference we list some properties of the functions $G$ and $H$.

**Lemma 2.4.18.** Under Conditions 2.4.1 and 2.4.2 the functions $G, H : (a,b) \to \mathbb{R}$ defined by $H(x) := I(x) - \gamma(x)F(x)$ and $G(x) := I(x) + \gamma(x)(1-F(x))$ have the following properties:

(a) $H$ and $G$ are continuous on $(a,b)$ and positive on $(a,b)$.

(b) $H$ is strictly increasing on $(a,b)$.

(c) $\lim_{x \to a} H(x) = 0$.

(d) $G$ is strictly decreasing on $(a,b)$.

(e) $\lim_{x \to b} G(x) = 0$.

(f) If $\gamma$ is absolutely continuous, then so are $H$ and $G$ and we have $H'(x) = -\gamma'(x)F(x)$ and $G'(x) = \gamma'(x)(1-F(x))$.

**Proof.** The assertion (a) has already been proven in the proofs of Propositions 2.4.10 and 2.4.16. To prove (b) and (d), fix $a < x < y < b$ and first observe that
2. Stein’s method for absolutely continuous univariate distributions

\[ I(y) - I(x) = \int_x^y \gamma(t)p(t)dt > \gamma(y)(F(y) - F(x)) \] \hspace{1cm} (2.25)

\[ I(y) - I(x) = \int_x^y \gamma(t)p(t)dt < \gamma(x)(F(y) - F(x)) \] \hspace{1cm} (2.26)

Thus, by (2.25) we obtain

\[ H(y) - H(x) = I(y) - I(x) + \gamma(x)F(x) - \gamma(y)F(y) \]
\[ > \gamma(y)(F(y) - F(x)) + \gamma(x)F(x) - \gamma(y)F(y) \]
\[ = F(x)(\gamma(x) - \gamma(y)) > 0 \]

since \( \gamma \) is strictly decreasing on \((a, b)\) and \( F \) is strictly positive there. Similarly, by (2.26) we have

\[ G(y) - G(x) = I(y) - I(x) + \gamma(y)(1 - F(y)) - \gamma(x)(1 - F(x)) < \gamma(x)(F(y) - F(x)) + \gamma(y)(1 - F(y)) - \gamma(x)(1 - F(x)) \]
\[ = (1 - F(y))(\gamma(y) - \gamma(x)) < 0 \]

since \( \gamma \) is strictly decreasing on \((a, b)\) and \( F(x) < 1 \).

Next, we prove (c) and (e). By (b) \( \lim_{x \to a} H(x) \) and \( \lim_{x \to b} G(x) \) exist finitely and are nonnegative by (a). Since \( \lim_{x \to a} I(x) = \lim_{x \to b} I(x) = 0 \), by the definitions of \( H \) and \( G \) we have \( \lim_{x \to a} (-\gamma(x)F(x)) \in [0, \infty) \) and \( \lim_{x \to b} \gamma(x)(1 - F(x)) \in [0, \infty) \). But \( -\gamma(x)F(x) \) is strictly negative for \( x \) near \( a \) and thus we also have \( \lim_{x \to a} H(x) \leq 0 \) and hence \( \lim_{x \to a} H(x) = 0 \). Analogously, \( \gamma(x)(1 - F(x)) \) is strictly negative for \( x \) near \( b \) which first implies \( \lim_{x \to b} G(x) \leq 0 \) and then \( \lim_{x \to b} G(x) = 0 \).

Of course, under the assumption of (f) both, \( H \) and \( G \), are also absolutely continuous because of their representation as a sum of products of such functions. Thus we only have to verify the formulas for their derivatives. We have

\[ H'(x) = I'(x) - \gamma'(x)F(x) - \gamma(x)p(x) = -\gamma'(x)F(x) \]

and so, because \( G = H + \gamma \),

\[ G'(x) = H'(x) + \gamma'(x) = (1 - F(x))\gamma'(x). \]

The assumption in part (f) of Lemma 2.4.18 will now be formulated as a condition.
2.4. Modification of Stein’s equation by exchangeable pairs

**Condition 2.4.19.** The function $\gamma$ fulfills all assumptions in Condition 2.4.2 and is additionally absolutely continuous on $(a,b)$.

If this condition is satisfied, then we can give another bound for $g'_h$ involving the derivative of $\gamma$.

**Proposition 2.4.20.** Assume Conditions 2.4.15 and 2.4.19. Then for any Lipschitz continuous test function $h : (a, b) \to \mathbb{R}$ we have for each $x \in (a, b)$

\[
g'_h(x) = \frac{1}{\eta(x)^2 p(x)} \int_a^b \int_x^b F(s)(1 - F(t))(h'(s)\gamma'(t) - \gamma'(s)h'(t))dsdt \tag{2.27}
\]

and

\[
|g'_h(x)| \leq \frac{2\|h'\|_\infty \|\gamma'\|_\infty \int_a^b F(s)ds \int_a^b (1 - F(t))dt}{\eta(x)^2 p(x)}. \tag{2.28}
\]

**Proof.** By Lemma 2.4.18 (c), (e) and (f) we have

\[
H(x) = -\int_a^x \gamma'(t)F(t)dt \quad \text{and} \quad G(x) = -\int_x^b \gamma'(t)(1 - F(t))dt. \tag{2.29}
\]

Plugging this into the representation

\[
g'_h(x) = \frac{1}{\eta(x)^2 p(x)} \left( G(x) \int_a^x F(s)h'(s)ds - H(x) \int_x^b (1 - F(s))h'(s)ds \right),
\]

which is just (2.23), yields the desired formula for $g'_h(x)$. This can now easily be bounded to obtain the claimed bound for $|g'_h(x)|$. \qed

**Corollary 2.4.21.** Assume Condition 2.4.15 and that $\gamma(x) = c(E[Z] - x)$ for some $c > 0$. Then we have for any Lipschitz continuous test function $h : (a, b) \to \mathbb{R}$ and each $x \in (a, b)$:

(a) $\|g_h\|_\infty \leq \frac{\|h'\|_\infty}{c}$

(b) $|g'_h(x)| \leq \frac{2\|h'\|_\infty H(x)G(x)}{\int_a^b \eta(x)p(x)} = 2\|h'\|_\infty \frac{\int_a^x F(s)ds \int_a^b (1 - F(t))dt}{\eta(x)(E[Z]F(x) - \int_a^x yp(y)dy)}$
2. Stein’s method for absolutely continuous univariate distributions

Proof. Claim (a) follows from Proposition 2.4.16 (a) and the observation that in this case we have

\[ I(x) = \int_a^x \gamma(y)p(y)dy = c \int_a^x (E[Z] - y)p(y)dy = c(E[Z]F(x) - \int_a^x yp(y)dy). \]

Part (b) easily follows from (2.28) and (2.29) but could also be derived directly from Proposition 2.4.16 (b).

Remark 2.4.22. (i) It is quite remarkable that in the case of normal approximation (via its classical Stein equation) the bound given in Corollary 2.4.21 (a) even improves on the best bound $2\|h'\|_{\infty}$ currently mentioned in the literature (see, e.g. [CGS11] or [CS05]). In fact, in this case $c = 1$ and thus our bound reduces to $\|h'\|_{\infty}$.

(ii) For concrete distributions the ratio appearing in the bound for $g_h'(x)$ may be bounded uniformly in $x$ by some constant which can sometimes also be computed explicitly. Nevertheless, for the case that $\gamma(x) = c(E[Z] - x)$ in [EV12] the authors give mild conditions for the existence of a finite constant $k$ such that $\|g_h'\|_{\infty} \leq k\|h'\|_{\infty}$ for any Lipschitz-continuous $h$. In practice, these conditions are usually met. However, there is no hope of estimating the constant $k$ by their method of proof. Thus, for concrete distributions and explicit constants it might therefore be useful to work with our bound from Corollary 2.4.21 (b).

(iii) For the normal distribution and also for the larger class of distributions discussed in [EL10], one also has a bound of the form $\|g_h''\|_{\infty} \leq C\|h'\|_{\infty}$ for some finite constant $C$ holding for each Lipschitz function $h$. As was shown by a universal example in [EV12], such a bound cannot be expected unless $a = -\infty$ and $b = \infty$, if one takes $\gamma(x) = c(E[Z] - x)$. This is why we will have to assume that $h'$ is also Lipschitz, for example by demanding that $h$ has two bounded derivatives. For the Stein solutions of the density approach, however, there are many examples of distributions, whose support is strictly included in $\mathbb{R}$ but for which such bounds are available (see, e.g., chapter 13 of [CGS11]).

Next, we will discuss, how we can express the density $p$ of $\mu$ in terms of $\gamma$ and $\eta$. This will be useful to bound the second derivative of $g_h$ in some special cases. Let $x_0$ be as in Condition 2.4.2. Since $\eta' = \gamma - \eta \psi$ and hence $\psi = \frac{\eta'}{\eta}$, we have
2.4. Modification of Stein’s equation by exchangeable pairs

\[ p(x) = p(x_0) \exp \left( \int_{x_0}^x \psi(t) \, dt \right) = p(x_0) \exp \left( \int_{x_0}^x \frac{\gamma(t)}{\eta(t)} \, dt \frac{\eta(x_0)}{\eta(x)} \right) \]

\[ = \frac{I(x_0)}{\eta(x)} \exp \left( \int_{x_0}^x \frac{\gamma(t)}{\eta(t)} \, dt \right). \tag{2.30} \]

Now, differentiating Stein’s equation (2.14), we obtain for \( h \) Lipschitz

\[ \eta(x) g''_h(x) + g'_h(x) \left( \eta'(x) + \gamma(x) \right) = h'(x) - \gamma'(x) g_h(x) =: h_2(x). \tag{2.31} \]

This means, that the function \( \tilde{g} := g'_h \) is a solution of the Stein equation corresponding to the test function \( h_2 \) for the distribution \( \mu \) which satisfies the Stein identity

\[ E \left[ \eta(Y) f'(Y) + (\eta'(Y) + \gamma(Y)) f(Y) \right] = 0, \]

where \( Y \sim \mu \). From (2.30) we know that a density \( \tilde{p} \) of \( \mu \) is given by

\[ \tilde{p}(x) = \frac{\tilde{K}}{\eta(x)} \exp \left( \int_{x_0}^x \frac{\eta'(t) + \gamma(t)}{\eta(t)} \, dt \right) = K \exp \left( \int_{x_0}^x \frac{\gamma(t)}{\eta(t)} \, dt \right), \tag{2.32} \]

where \( \tilde{K}, K > 0 \) are suitable normalizing constants. Thus, if we have bounds for the first derivative of the Stein solutions for the distribution \( \mu \) and for Lipschitz functions \( h \), then we obtain from this observation bounds on \( g''_h \) for \( h \) such that \( h_2 \) is Lipschitz (if \( \gamma(x) = c(E[Z] - x) \), this essentially means that \( h' \) must be Lipschitz). Of course, it has to be verified, that \( \mu(h_2) = 0 \) in order to use these bounds. But this will follow immediately, if one can show that \( \tilde{g} \) belongs to a class of functions for which the Stein identity for \( \mu \) is valid. If \( \mu = \mu_{\alpha, \beta} \) belongs to the family of Beta distributions, then so does \( \tilde{\mu} = \mu_{\alpha + 1, \beta + 1} \), since in this case, with obvious notation, \( \eta_{\alpha, \beta}(x) := \eta(x) = 1 - x^2 \) and \( \gamma_{\alpha, \beta}(x) = -[(\alpha + \beta + 2)x + \alpha - \beta] \). Hence, we have \( \eta'(x) + \gamma_{\alpha, \beta}(x) = \gamma_{\alpha + 1, \beta + 1}(x) \). Since in this case, for \( h \) such that \( h' \) is Lipschitz, we have \( g'_h = \tilde{g} \in K_{\alpha + 1, \beta + 1} \) (see Section 2.5), we can apply the proven bounds for the first derivative of the Stein solution corresponding to the distribution \( \mu_{\alpha + 1, \beta + 1} \) to the function \( g'_h \).

It is a well-known fact from measure and integration theory, that if a function \( f : \mathbb{R} \to \mathbb{R} \) is differentiable with bounded derivative, then \( f' \) is Lebesgue-integrable on each compact interval \([c, d]\) and \( f(d) - f(c) = \int_c^d f'(x) \, dx \). By monotonicity of the integral, this implies that \( f \) is Lipschitz-continuous with Lipschitz constant \( \|f'\|_{\infty} \). This yields the following result.

**Lemma 2.4.23.** Let \( h : \mathbb{R} \to \mathbb{R} \) be twice differentiable with bounded second derivative. Then \( h' \) is Lipschitz with Lipschitz constant \( \|h''\|_{\infty} \).
2. Stein’s method for absolutely continuous univariate distributions

In the following we are going to readdress the problem of support as discussed in Section 2.2 and show how Stein’s equation and its solution may be extended. To this end, we suppose that the functions \( \gamma \) and \( \eta \) both have a natural extension to all of \( \mathbb{R} \), for example that they are given by some “analytical expression” on \((a, b)\) which still makes sense on \( \mathbb{R} \setminus (a, b) \). We will need the following condition.

**CONDITION 2.4.24.** The functions \( \gamma \) and \( \eta \) are defined as before on \((a, b)\) and may be extended on \( \mathbb{R} \) such that the following properties are satisfied:

(i) On \((a, b)\) the function \( \gamma \) has all the properties listed in Condition 2.4.1 and is continuous and strictly decreasing on \( \mathbb{R} \).

(ii) The function \( \eta \) is given by (2.18) for \( x \in (a, b) \) and is absolutely continuous on every compact sub-interval \([c, d] \subseteq \mathbb{R}\). Furthermore, \( \eta(x) \neq 0 \) for all \( x \in \mathbb{R} \setminus \{a, b\} \).

Let \( h : \mathbb{R} \to \mathbb{R} \) be a given Borel-measurable test function with \( E[|h(Z)|] < \infty \). Then we have the following Stein equation, valid now for all \( x \in \mathbb{R} \):

\[
\eta(x)g'(x) + \gamma(x)g(x) = h(x) - \mu(h) =: \tilde{h}(x) \tag{2.33}
\]

We already know, how to solve (2.33) for \( x \in (a, b) \). So now, we will assume that at least one of \( a \) and \( b \) is finite and try to solve the equation outside \((a, b)\). Furthermore, we will discuss conditions that ensure that the composed solution \( g_b \) behaves nicely at the boundary points \( a \) and/or \( b \). We will henceforth assume Condition 2.4.24.

For \( x \neq a, b \) equation (2.33) is clearly equivalent to

\[
g'(x) = -\frac{\gamma(x)}{\eta(x)} g(x) + \frac{\tilde{h}(x)}{\eta(x)}. \tag{2.34}
\]

Let us assume that both \( a > -\infty \) and \( b < \infty \) (the other cases are of course included) and let \( F_i \) be any primitive function of \( \frac{2}{\eta} \) on \((-\infty, a)\). Such a function exists by continuity and is hence continuously differentiable. Fix \( d \in (-\infty, a) \). The solution \( \varphi_1 \) of the homogeneous equation corresponding to (2.34) with \( \varphi_1(d) = 1 \) is given by

\[
\varphi_1(x) = \exp \left( - \int_d^x \frac{\gamma(t)}{\eta(t)} \, dt \right) = \exp(F_i(d)) \exp(-F_i(x)).
\]

By the method of variation of the constant, the solution \( g \) of (2.34) on \((-\infty, a)\) with \( g(d) = y_0 \) is given by

46
2.4. Modification of Stein’s equation by exchangeable pairs

\[ g(x) = \left( y_0 + \int_d^x \frac{\tilde{h}(t)}{\eta(t)} \exp\left( \int_d^t \frac{\gamma(s)}{\eta(s)} \, ds \right) \, dt \right) \exp(F_i(d)) \exp(-F_i(x)) \]

\[ = \left( y_0 + \int_d^x \frac{\tilde{h}(t)}{\eta(t)} \exp(F_i(t)) \, dt \exp(-F_i(d)) \right) \exp(F_i(d)) \exp(-F_i(x)) \]

\[ = y_0 \exp(F_i(d)) \exp(-F_i(x)) + \exp(-F_i(x)) \int_d^x \frac{\tilde{h}(t)}{\eta(t)} \exp(F_i(t)) \, dt . \]

First choosing \( y_0 = g(d) = 0 \) and then letting \( d \nearrow a \) yields

\[ g_h(x) := \exp(-F_i(x)) \int_a^x \frac{\tilde{h}(t)}{\eta(t)} \exp(F_i(t)) \, dt , \quad (2.35) \]

if the integral still exists. Note that this property does not depend on the particular choice of the primitive function \( F_i \). For a fixed primitive function \( F_i \) of \( \tilde{\eta} \) on \((-\infty, a)\) we define the function

\[ q_i := \frac{\exp(F_i)}{\eta} . \]

Analogously, for a given primitive function \( F_r \) of \( \tilde{\eta} \) on \((b, \infty)\) we define the function

\[ q_r := \frac{\exp(F_r)}{\eta} . \]

Note that inside the interval \((a, b)\) we have that \( \log(\eta \eta) \) is a primitive of \( \tilde{\eta} \) and hence \( q_i \) plays the role of \( p \) on \((-\infty, a)\) (and similarly for \( q_r \)). As we have observed, we will need the following Condition:

**Condition 2.4.25.** There exist primitive functions \( F_i : (-\infty, a) \to \mathbb{R} \) and \( F_r : (b, \infty) \to \mathbb{R} \) of \( \tilde{\eta} \) on \((-\infty, a)\) and on \((b, \infty)\), respectively, such that for each \( x \in (-\infty, a) \) the function \( q_i \) is integrable over \([x, a)\) and for each \( y \in (b, \infty) \) the function \( q_r \) is integrable over \([b, y)\). We may thus define functions \( Q_l : (-\infty, a) \to \mathbb{R} \) by \( Q_l(x) := \int_a^x q_i(t) \, dt \) and \( Q_r : (b, \infty) \to \mathbb{R} \) by \( Q_r(y) := \int_b^y q_r(t) \, dt \).

Similar to the above derivation, for \( x \in (b, \infty) \) we arrive at the definition

\[ g_h(x) := \exp(-F_r(x)) \int_b^x \frac{\tilde{h}(t)}{\eta(t)} \exp(F_r(t)) \, dt = \exp(-F_r(x)) \int_b^x \tilde{h}(t) q_r(t) \, dt , \quad (2.36) \]

if this integral exists. So, in the following we will always implicitly assume that also \( \int_y^x |\tilde{h}(t) q_r(t)| \, dt < \infty \) and \( \int_x^a |\tilde{h}(t) q_i(t)| \, dt < \infty \) both hold for each \( x \in (-\infty, a) \) and each \( y \in (b, \infty) \).
Remark 2.4.26. Note that the definition of the solution \( g_h \) does not depend on the choice of the primitive functions \( F_l \) and \( F_r \) since two such functions may only differ by an additive constant.

Now we investigate under what conditions the function \( g_h \), up to now only defined on \( \mathbb{R} \setminus \{a, b\} \), has a continuous extension on \( \mathbb{R} \). Of course \( g_h \) is absolutely continuous on every compact interval \([c, d]\) which is included in either \((-\infty, a)\), \((a, b)\) or \((b, \infty)\). By the properties of the function \( I = pq \) on \((a, b)\), from Proposition 2.4.5, the continuity of \( \gamma \) and by de l'Hôpital's rule (see Theorem A.1.5) we have

\[
\lim_{x \searrow a} g_h(x) = \lim_{x \searrow a} \frac{\int_a^x \bar{h}(t)p(t)dt}{I(x)} = \lim_{x \searrow a} \frac{\bar{h}(x)p(x)}{\gamma(x)p(x)} = \lim_{x \searrow a} \frac{\gamma(x)}{\gamma(a)} \lim_{x \searrow a} \bar{h}(x)
\]

(2.37)

and

\[
\lim_{x \nearrow b} g_h(x) = \lim_{x \nearrow b} \frac{\int_a^x \bar{h}(t)p(t)dt}{I(x)} = \lim_{x \nearrow b} \frac{\bar{h}(x)p(x)}{\gamma(x)p(x)} = \lim_{x \nearrow b} \frac{\gamma(x)}{\gamma(b)} \lim_{x \nearrow b} \bar{h}(x),
\]

(2.38)

if these limits exist.

So, the limits in (2.37) and in (2.38) do exist and coincide if \( h \) is continuous at \( a \) and at \( b \).

To deal with the limits \( \lim_{x \searrow a} g_h(x) \) and \( \lim_{x \nearrow b} g_h(x) \) we first formulate a condition which will usually be satisfied in practice.

Condition 2.4.27. The functions \( F_l \) and \( F_r \) satisfy \( \lim_{x \searrow a} F_l(x) = \pm \infty \) and \( \lim_{x \nearrow b} F_r(x) = \pm \infty \).

Again, the validity of this condition does not depend on the choice of the functions \( F_l \) and \( F_r \). By Condition 2.4.27 we may again apply de l'Hôpital's rule to compute

\[
\lim_{x \searrow a} g_h(x) = \lim_{x \searrow a} \frac{\int_a^x \bar{h}(t)q_l(t)dt}{\exp(F_l(x))} = \lim_{x \searrow a} \frac{\bar{h}(x)q_l(x)}{\exp(F_l(x))F'_l(x)} = \lim_{x \searrow a} \frac{\gamma(x)}{\gamma(a)} \lim_{x \searrow a} \bar{h}(x)
\]

(2.39)

and

\[
\lim_{x \nearrow b} g_h(x) = \lim_{x \nearrow b} \frac{\int_b^x \bar{h}(t)q_r(t)dt}{\exp(F_r(x))} = \lim_{x \nearrow b} \frac{\bar{h}(x)q_r(x)}{\exp(F_r(x))F'_r(x)} = \lim_{x \nearrow b} \frac{\gamma(x)}{\gamma(b)} \lim_{x \nearrow b} \bar{h}(x),
\]

(2.40)
2.4. Modification of Stein’s equation by exchangeable pairs

if these limits exist. Again, the continuity of \( h \) at \( a \) and \( b \) is sufficient for this to hold. We can thus conclude the following proposition.

**Proposition 2.4.28.** Assume Conditions 2.4.24, 2.4.25 and 2.4.27 and let \( h : \mathbb{R} \to \mathbb{R} \) be a Borel-measurable test function satisfying the above integrability conditions and being continuous at \( a \) and \( b \). Then the Stein solution \( g_h \) as defined above may be extended to a continuous function on \( \mathbb{R} \) by letting

\[
g_h(a) := \frac{h(a) - \mu(h)}{\gamma(a)} \quad \text{and} \quad g_h(b) := \frac{h(b) - \mu(h)}{\gamma(b)}.
\]

Next, we want to present bounds on \( g_h(x) \) and \( g_h'(x) \) for \( x \notin (a, b) \). But first we will show that our conditions already imply that \( \eta(x) < 0 \) if \( x < a \) or \( x > b \). Since \( F_1'(x) = \frac{\eta(x)}{\eta(x)} \) is then negative on \((-\infty, a)\) this also ensures that \( \lim_{x \nearrow a} F_1(x) = -\infty \). Similarly, \( \lim_{x \searrow b} F_r(x) = -\infty \).

**Proposition 2.4.29.** Assume Conditions 2.4.24, 2.4.25 and 2.4.27. Then the functions \( \eta \), \( F_1 \) and \( F_r \) have the following properties:

(a) For all \( x \in \mathbb{R} \setminus (a, b) \) we have \( \eta(x) < 0 \). Especially, by continuity we have \( \eta(a) = \eta(b) = 0 \).

(b) We have \( \lim_{x \nearrow a} F_1(x) = \lim_{x \searrow b} F_r(x) = -\infty \).

**Proof.** To prove (a), first note, that by Condition 2.4.24 \( \eta \) has no sign changes on \((-\infty, a)\). Suppose contrarily to the assertion, that \( \eta(x) > 0 \) for all \( x \in (-\infty, a) \). Then, the function \( q_1 \) is also positive and hence \( 0 \leq \int_a^x q_1(t) \, dt = -Q_1(x) \) for each \( x \in (-\infty, a) \). By Conditions 2.4.25 and 2.4.27 we may apply de l’Hôpital’s rule to conclude

\[
0 \leq \lim_{x \nearrow a} \frac{-Q_1(x)}{\exp(F_1(x))} = \lim_{x \nearrow a} \frac{-q_1(x)}{\frac{\gamma(x)}{\eta(x)} \exp(F_1(x))} = \lim_{x \nearrow a} \frac{-q_1(x)}{\gamma(x)q_1(x)} = -\frac{1}{\gamma(a)} < 0,
\]

by Condition 2.4.24. This is a contradiction and hence we must have \( \eta(x) < 0 \) for \( x \in (-\infty, a) \). Similarly, one shows that also \( \eta(x) < 0 \) for \( x \in (b, \infty) \).

To prove (b), note that \( F_1'(x) = \frac{\gamma(x)}{\eta(x)} < 0 \) for \( x < a \). By Condition 2.4.27 this necessarily implies that \( \lim_{x \nearrow a} F_1(x) = -\infty \). Analogously, we have \( F_r'(x) > 0 \) for \( x > b \) (since \( \gamma(x) < 0 \) there) which by Condition 2.4.24 implies that \( \lim_{x \searrow b} F_r(x) = -\infty \). \( \square \)

49
Proposition 2.4.29 particularly implies the conclusion of Proposition 2.4.9 and hence makes Condition 2.4.7 redundant, at least as far as this proposition is concerned.

In order to get general bounds, we will need yet another condition on the functions \( F_l \) and \( F_r \).

**CONDITION 2.4.30.** The functions \( F_l \) and \( F_r \) satisfy
\[
\lim_{x \to -\infty} F_l(x) = \lim_{x \to \infty} F_r(x) = +\infty.
\]

Now we are in the position to prove bounds on \( g_h \) for bounded, Borel-measurable functions \( h \).

**PROPOSITION 2.4.31.** Assume Conditions 2.4.24, 2.4.25, 2.4.27 and 2.4.30 and let \( m \) be a median for \( \mu \). Then, for any bounded and Borel-measurable test function \( h : \mathbb{R} \to \mathbb{R} \) we have
\[
\|g_h\|_\infty \leq \|h - \mu(h)\|_\infty \max\left(\frac{1}{2I(m)}, \frac{1}{\gamma(a)}, \frac{-1}{\gamma(b)}\right).
\]

**Proof.** The bound on \( |g_h(x)| \) for \( x \in (a, b) \) has already been proved in Proposition 2.4.10. Let \( x \in (-\infty, a) \). Then we have by the negativity of \( q_l \) which follows from Proposition 2.4.29:
\[
|g_h(x)| = \left| \int_a^x \hat{h}(t)q_l(t)dt \right| \leq \|\hat{h}\|_\infty \left| \int_a^x q_l(t)dt \right| \leq \|\hat{h}\|_\infty \frac{\int_a^x q_l(t)dt}{\exp(F_l(x))} = \|\hat{h}\|_\infty \frac{Q_l(x)}{\exp(F_l(x))}.
\]

We want to show, that this is bounded from above by \( \frac{\|\hat{h}\|_\infty}{\gamma(a)} \). To this end, we define the function \( D(x) := \frac{\exp(F_l(x))}{\gamma(a)} - Q_l(x), \ x \in (-\infty, a) \), and show that \( D(x) > 0 \). By Condition 2.4.25 and Proposition 2.4.29 we have \( \lim_{x \to a} D(x) = 0 \) and furthermore
\[
D'(x) = \frac{\gamma(x)}{\gamma(a)} \exp(F_l(x)) - q_l(x) = q_l(x)\left(\frac{\gamma(x)}{\gamma(a)} - 1\right) < 0,
\]

since \( q_l(x) < 0 \) and \( \gamma(x) > \gamma(a) \) by Condition 2.4.24. Thus, \( D \) is strictly decreasing and hence \( D(x) > 0 \) for each \( x \in (-\infty, a) \). This proves the desired bound for \( x \in (-\infty, a) \). Similarly one proves that \( |g_h(x)| \leq -\frac{\|\hat{h}\|_\infty}{\gamma(b)} \) for each \( x \in (b, \infty) \).
2.4. Modification of Stein’s equation by exchangeable pairs

Before turning to Lipschitz test functions, we discuss properties of the functions \( \exp(F_t) \) and \( \exp(F_r) \), respectively. In particular, we will show, that they correspond to the function \( I \) on \((a, b)\) and have similar integral representations.

**Proposition 2.4.32.** We define the functions \( I_l : (-\infty, a) \to \mathbb{R} \) and \( I_r : (b, \infty) \to \mathbb{R} \) by \( I_l(x) := \exp(F_l(x)) \) and \( I_r(x) := \exp(F_r(x)) \). Then the following representations hold:

(a) For each \( x \in (-\infty, a) \) we have \( I_l(x) = \int_a^x \gamma(t)q_l(t)dt \).

(b) For each \( x \in (b, \infty) \) we have \( I_r(x) = \int_b^x \gamma(t)q_r(t)dt \).

In particular, these integrals exist.

**Proof.** We only prove (a). Let \( x \in (-\infty, a) \) and let \((a_n)_{n \in \mathbb{N}}\) be any sequence in \((-\infty, a)\) with \( \lim_{n \to \infty} a_n = a \). Then, for each \( n \in \mathbb{N} \)

\[
\int_{a_n}^x \gamma(t)q_l(t)dt = \int_{a_n}^x \frac{\gamma(t)}{\eta(t)} \exp(F_l(t))dt = \int_{a_n}^x F_l'(t) \exp(F_l(t))dt = \int_{F_l(a_n)}^{F_l(x)} e^s ds = \exp(F_l(x)) - \exp(F_l(a_n)) \xrightarrow{n \to \infty} \exp(F_l(x)),
\]

by Proposition 2.4.29. Hence \( \int_a^x \gamma(t)q_l(t)dt \) exists and equals \( I_l(x) \).

Next, we will state a lemma, which replaces Lemma 2.4.13 outside the support \((a, b)\).

**Lemma 2.4.33.** (a) For each \( x \in (-\infty, a) \) we have \( \int_x^a Q_l(t)dt = -xQ_l(x) + \int_x^a t q_l(t)dt \) and \( I_l(x) < \gamma(x)Q_l(x) \).

(b) For each \( x \in (b, \infty) \) we have \( \int_b^x Q_r(t)dt = xQ_r(x) - \int_b^x t q_r(t)dt \) and \( I_r(x) < \gamma(x)Q_r(x) \).

**Proof.** By Fubini's theorem we have

\[
\int_x^a Q_l(t)dt = \int_x^a \left( \int_t^a q_l(s)ds \right) dt = -\int_x^a \left( \int_t^a q_l(s)ds \right) dt = -\int_x^a \left( \int_x^s 1dt \right) q_l(s)ds = -\int_x^a (s-x)q_l(s)ds = -xQ_l(x) + \int_x^a s q_l(s)ds,
\]
2. Stein’s method for absolutely continuous univariate distributions

which proves the first part of (a). The second claim of (a) follows from (a), the positivity of $-q_l$ on $(-\infty, a)$ and from the monotonicity of $\gamma$:

$$I_l(x) = \int_x^a \gamma(t)q_l(t)dt = \int_x^a \gamma(t)(-q_l(t))dt$$

$$< \gamma(x) \int_x^a (-q_l(t))dt = \gamma(x) \int_x^a q_l(t)dt$$

$$= \gamma(x)Q_l(x)$$

The proof of (b) is similar but easier, and is therefore omitted. \qed

**Lemma 2.4.34.** Let $h : \mathbb{R} \to \mathbb{R}$ be Lipschitz-continuous.

(a) For each $x \in (-\infty, a)$ we have

$$h(x) - \mu(h) = -\int_x^b (1 - F(s))h'(s)ds = -\int_x^a h'(s)ds - \int_a^b (1 - F(s))h'(s)ds$$

and

$$g_h(x) = \frac{1}{I_l(x)} \left( -\int_x^a (Q_l(x) - Q_l(s))h'(s)ds - Q_l(x) \int_a^b (1 - F(s))h'(s)ds \right)$$

$$= \frac{1}{I_l(x)} \left( \int_x^a Q_l(s)h'(s)ds - Q_l(x) \int_x^a (1 - F(s))h'(s)ds \right).$$

(b) For each $x \in (b, \infty)$ we have $h(x) - \mu(h) = \int_a^x F(s)h'(s)ds$ and

$$g_h(x) = \frac{1}{I_r(x)} \left( Q_r(x) \int_a^b F(s)h'(s)ds + \int_a^x h'(s)(Q_r(x) - Q_r(s))ds \right)$$

$$= \frac{1}{I_r(x)} \left( Q_r(x) \int_a^x F(s)h'(s)ds - \int_b^x Q_r(s)h'(s)ds \right).$$

*Proof.* We only prove (a) since the proof of (b) is very similar. The first claim follows from Lemma 2.4.14 (a) since $F(s) = 0$ for $s < a$ and $F(s) = 1$ for $x \geq b$. The second claim follows from the first one and from Fubini’s theorem by
For each \( h \)

Then, for any Lipschitz-continuous test function \( h \):

**Proposition 2.4.35.** Assume Conditions 2.4.15, 2.4.24, 2.4.25, 2.4.27 and 2.4.30. Then, for any Lipschitz-continuous test function \( h : \mathbb{R} \to \mathbb{R} \) the following bounds hold true:

(a) For each \( x \in (-\infty, a) \) we have \( |g_h(x)| \leq \|h'\|_\infty \frac{Q_t(x)E[Z] - \int_x^t q_s(s)ds}{I_t(x)} \).

(b) For each \( x \in (b, \infty) \) we have \( |g_h(x)| \leq \|h'\|_\infty \frac{Q_t(x)E[Z] - \int_x^t q_s(s)ds}{I_t(x)} \).

(c) For each \( x \in (-\infty, a) \) we have

\[
|g_h(x)| \leq \frac{\|h'\|_\infty}{-\eta(x)I_t(x)} \left( \gamma(x) \left( -xQ_t(x) + \int_a^x t q_l(t)dt \right) + (E[Z] - x) \left( Q_t(x)\gamma(x) - I_t(x) \right) \right).
\]
2. Stein’s method for absolutely continuous univariate distributions

(d) For each \( x \in (b, \infty) \) we have

\[
|g'_h(x)| \leq \frac{\|h'\|_\infty}{-\eta(x)I_l(x)} \left( \gamma(x) \left( xQ_t(x) - \int_b^x tq_r(t) dt \right) + (x - E[Z]) \left( \gamma(x)Q_r(x) - I_r(x) \right) \right).
\]

Proof. We only prove (a) and (c), since the proofs of (b) and (d) are similar. To prove (a), we observe that by Lemma 2.4.34 we have

\[
g_h(x) = \frac{1}{I_l(x)} \left( - \int_x^a (Q_t(x) - Q_t(s)) h'(s) ds - Q_t(x) \int_a^b (1 - F(s)) h'(s) ds \right).
\]

Since \( Q_t \) is decreasing \( (Q'_t = q_t < 0) \) and positive on \( (-\infty, a) \) this implies

\[
|g_h(x)| \leq \frac{\|h'\|_\infty}{I_l(x)} \left( \int_x^a (Q_t(x) - Q_t(s)) ds + Q_t(x) \int_a^b (1 - F(s)) ds \right).
\]

By Lemma 2.4.13 and Lemma 2.4.33 (a) the right hand side equals

\[
\frac{\|h'\|_\infty}{I_l(x)} \left( (a - x)Q_l(x) + xQ_l(x) - \int_a^x sq_l(s) ds + Q_l(x) \left( E[Z] - a \right) \right)
\]

\[
= \frac{\|h'\|_\infty}{I_l(x)} \left( Q_l(x)E[Z] - \int_a^x sq_l(s) ds \right).
\]

which is the claimed bound. Now, we turn to the proof of (c). By Stein’s equation (2.33) and Lemma 2.4.34 (a) we have for each \( x \in (-\infty, a) \):

\[
g'_h(x) = \frac{\tilde{h}(x)}{\eta(x)} - \frac{\gamma(x)}{\eta(x)} g_h(x) = -\frac{1}{\eta(x)} \int_x^b (1 - F(s)) h'(s) ds
\]

\[
-\frac{\gamma(x)}{\eta(x)I_l(x)} \left( \int_x^a Q_t(s) h'(s) ds - Q_t(x) \int_x^b (1 - F(s)) h'(s) ds \right)
\]

\[
= -\frac{1}{\eta(x)I_l(x)} \left( \gamma(x) \int_x^a Q_t(s) h'(s) ds + \int_x^b (1 - F(s)) h'(s) ds \left( I_l(x) - Q_t(x) \gamma(x) \right) \right)
\]

By Lemma 2.4.13 (a) we have
2.4. Modification of Stein’s equation by exchangeable pairs

\[
\int_x^b (1 - F(s)) \, ds = a - x + \int_x^b (1 - F(s)) \, ds = a - x + E[Z] - a = E[Z] - x.
\]

Hence, by Lemma 2.4.33 (a) this implies

\[
\left| g_h'(x) \right| \leq \frac{\| h' \|_{\infty}}{-\eta(x) I_l(x)} \left( \gamma(x) \left( \int_x^a Q_l(s) \, ds \right) + \int_x^b (1 - F(s)) \, ds \left( Q_l(x) \gamma(x) - I_l(x) \right) \right)
\]

\[
= \frac{\| h' \|_{\infty}}{-\eta(x) I_l(x)} \left( \gamma(x) \left( -x Q_l(x) + \int_a^x t q_l(t) \, dt \right) + (E[Z] - x) \left( Q_l(x) \gamma(x) - I_l(x) \right) \right)
\]

which was to be shown.

Corollary 2.4.36. Assume Conditions 2.4.15, 2.4.24, 2.4.25, 2.4.27 and 2.4.30 and that \( \gamma(x) = c(E[Z] - x) \) for some \( c > 0 \). Then, for any Lipschitz-continuous test function \( h : \mathbb{R} \to \mathbb{R} \) one has the following bounds:

(a) \( \| g_h \|_{\infty} \leq \frac{\| h' \|_{\infty}}{c} \)

(b) For each \( x \in (-\infty, a) \) we have

\[
\left| g_h'(x) \right| \leq 2 \| h' \|_{\infty} \frac{\gamma(x) \left( -x Q_l(x) + \int_a^x t q_l(t) \, dt \right)}{-\eta(x) I_l(x)}.
\]

(c) For each \( x \in (b, \infty) \) we have

\[
\left| g_h'(x) \right| \leq 2 \| h' \|_{\infty} \frac{\gamma(x) \left( x Q_r(x) - \int_b^x t q_r(t) \, dt \right)}{-\eta(x) I_l(x)}.
\]

Proof. In this case, we have the relations

\[
I_l(x) = \int_a^x (E[Z] - t) q_l(t) \, dt = c \left( Q_l(x) E[Z] - \int_a^x t q_l(t) \, dt \right), \quad (2.41)
\]

\[
I_r(x) = \int_b^x (E[Z] - t) q_r(t) \, dt = c \left( Q_r(x) E[Z] - \int_b^x t q_r(t) \, dt \right).
\]
2. Stein’s method for absolutely continuous univariate distributions

These together with Proposition 2.4.35 (a), (b) and Corollary 2.4.21 immediately imply (a). As to (b), by (2.41) we have

\[
(E[Z] - x) (Q_l(x) \gamma(x) - I_l(x))
\]

\[
= (E[Z] - x) \left( cE[Z]Q_l(x) - cxQ_l(x) - c(Q_l(x)E[Z] - \int_a^x t q_l(t) dt) \right)
\]

\[
= c(E[Z] - x)(-xQ_l(x) + \int_a^x t q_l(t) dt)
\]

\[
= \gamma(x)(-xQ_l(x) + \int_a^x t q_l(t) dt).
\]

This and Proposition 2.4.35 (c) imply claim (b). Assertion (c) may be proved similarly.

Next, we will give a short account of the properties of the solutions \(g_z\) corresponding to the test functions \(h_z := 1_{(-\infty, z]}\), \(z \in \mathbb{R}\), yielding the Kolmogorov distance. In general, these functions are only of interest for \(z \in (a, b)\) since for any real-valued random variable \(W\) and any \(z \leq a\) we have

\[
\left| P(W \leq z) - P(Z \leq z) \right| = P(W \leq z)
\]

and for any \(z \geq b\) we have

\[
\left| P(W \leq z) - P(Z \leq z) \right| = 1 - P(W \leq z) = P(W > z).
\]

These probabilities have nothing to do with the distribution \(\mu\) of \(Z\) and hence, one should be able to bound them as well directly for a given \(W\). Thus, we will focus on \(z \in (a, b)\).

**Proposition 2.4.37.** Let \(z \in (a, b)\) be given. Then, under the Conditions 2.4.1, 2.4.24, 2.4.25 and 2.4.27 we have:

\[
g_z(x) = \begin{cases} 
(1-F(z))Q_l(x) & \text{, } x < a \\
1-F(z) & \text{, } x = a \\
\frac{\gamma(a)}{F(x)(1-F(z))} & \text{, } a < x \leq z \\
\frac{F(z)(1-F(x))}{I_l(x)} & \text{, } z < x < b \\
-\frac{F(z)}{\gamma(b)} & \text{, } x = b \\
-\frac{F(z)Q_r(x)}{I_r(x)} & \text{, } x > b
\end{cases}
\]
2.4. Modification of Stein’s equation by exchangeable pairs

Furthermore, one has the bounds \( \|g_z\|_\infty = \frac{F(z)(1-F(z))}{I(z)} \) and also \( \sup_{x \in (a,b)} \|g_z\|_\infty < \infty \). The functions \( g_z \) are absolutely continuous on every compact sub-interval of \( \mathbb{R} \) and

\[
g'_z(x) = \begin{cases} \frac{-(1-F(z))q(x)(\gamma(x)Q_x(x)-I_x(x))}{I_x(x)^2}, & x < a \\ \frac{(1-F(z)p(x)(I_x(x)-\gamma(x)F(x))}{I_x(x)^2}, & a < x < z \\ \frac{-F(z)p(x)(I_x(x)+\gamma(x)(1-F(x))}{I_x(x)^2}, & z < x < b \\ \frac{-F(z)p(x)(I_x(x)-\gamma(x)Q_x(x))}{I_x(x)^2}, & x > b \end{cases}
\]

Proof. For any \( x \in (a, b) \) we have:

\[
g_z(x) = \frac{1}{I(x)} \int_a^x (1_{(-\infty,z]}(t) - P(Z \leq z))p(t)dt
\]

\[
= \frac{1}{I(x)} \left( \int_a^x p(t)dt - F(z)F(x) \right)
\]

\[
= \frac{F(x \land z) - F(z)F(x)}{I(x)},
\]

proving the desired representation of \( g_z \) inside the interval \((a, b)\). Now, for \( x \in (a, b) \) let \( M(x) := \frac{F(x)}{I(x)} \) and \( N(x) := \frac{1-F(x)}{I(x)} \). Then we have

\[
M'(x) = \frac{p(x)I(x) - p(x)\gamma(x)F(x)}{I(x)^2} = \frac{p(x)}{I(x)^2} \left( I(x) - \gamma(x)F(x) \right)
\]

\[
= \frac{p(x)}{I(x)^2} H(x) > 0
\]

and

\[
N'(x) = \frac{-p(x)I(x) - p(x)\gamma(x)(1-F(x))}{I(x)^2} = \frac{-p(x)}{I(x)^2} \left( I(x) + (1-F(x))\gamma(x) \right)
\]

\[
= \frac{-p(x)}{I(x)^2} G(x) < 0,
\]

by Lemma 2.4.18. Thus, \( M \) is strictly increasing and \( N \) is strictly decreasing on \((a, b)\). Since \( g_z(x) = (1-F(z))M(x) \) for \( x \in (a, z) \) and \( g_z(x) = F(z)N(x) \) for \( x \in (z, b) \), this implies, that \( \sup_{x \in (a,b)} g_z(x) = g_z(z) = \frac{F(z)(1-F(z))}{I(z)} \). It also implies the claimed representation of \( g'_z(x) \) for \( x \in (a, b) \setminus \{z\} \). Furthermore, by de l’Hôpital’s rule, we have
2. Stein’s method for absolutely continuous univariate distributions

\[
\lim_{x \searrow a} g_z(x) = (1 - F(z)) \lim_{x \searrow a} M(x) = (1 - F(z)) \lim_{x \searrow a} \frac{p(x)}{p(x)\gamma(x)} = \lim_{x \searrow a} \frac{(1 - F(z))}{\gamma(x)} = \frac{(1 - F(z))}{\gamma(a)}
\]

and

\[
\lim_{x \nearrow b} g_z(x) = F(z) \lim_{x \nearrow b} N(x) = F(z) \lim_{x \nearrow b} \frac{-p(x)}{p(x)\gamma(x)} = \lim_{x \nearrow b} \frac{-F(z)}{\gamma(x)} = \frac{-F(z)}{\gamma(b)}.
\]

Note, that these limits could also be derived from Proposition 2.4.28.

Next, consider \(x \in (-\infty, a)\). For such an \(x\) we have

\[
g_z(x) = \frac{1}{I_l(x)} \int_a^x \left(1 - F(Z \leq z)\right) q_l(t) dt = \frac{(1 - F(z))Q_l(x)}{I_l(x)}.
\]

Moreover we have

\[
\frac{d}{dx} \frac{Q_l(x)}{I_l(x)} = \frac{q_l(x)I_l(x) - \gamma(x)q_l(x)Q_l(x)}{I_l(x)^2} = \frac{-q_l(x)}{I_l(x)^2} \left(\gamma(x)Q_l(x) - I_l(x)\right) > 0
\]

by Lemma 2.4.33. Thus, \(g_z\) is increasing on \((-\infty, a)\) and hence, again by de l’Hôpital’s rule,

\[
\sup_{x \in (-\infty, a)} |g_z(x)| = \sup_{x \in (-\infty, a)} g_z(x) = (1 - F(z)) \lim_{x \nearrow a} \frac{Q_l(x)}{I_l(x)} = (1 - F(z)) \lim_{x \nearrow a} \frac{q_l(x)}{q_l(x)\gamma(x)} = (1 - F(z)) \lim_{x \nearrow a} \frac{1}{\gamma(x)} = \frac{1 - F(z)}{\gamma(a)}.
\]

Since \(g_z'(x) = (1 - F(z)) \frac{d}{dx} \frac{Q_l(x)}{I_l(x)}\) we have also derived the desired formula for \(g_z'(x)\) for \(x \in (-\infty, a)\). The calculations for \(x \in (b, \infty)\) are completely analogous.
2.4. Modification of Stein’s equation by exchangeable pairs

and therefore omitted. From our computations we can already infer, that \( \|g_z\|_\infty = \frac{F(z)(1-F(z))}{I(z)} \). So, it remains to show that this quantity is bounded in \( z \in (a, b) \). Since it is a continuous function of \( z \), we only have to show that it has finite limits on the edge of the interval \( (a, b) \). But, of course,

\[
\lim_{z \searrow a} \frac{F(z)(1-F(z))}{I(z)} = \lim_{z \searrow a} \frac{F(z)}{I(z)} = \lim_{z \searrow a} M(z) = \frac{1}{\gamma(a)}
\]

and, similarly, \( \lim_{z \nearrow b} \frac{F(z)(1-F(z))}{I(z)} = \lim_{z \nearrow b} N(z) = -\frac{1}{\gamma(b)} \). This concludes the proof. \( \square \)

**Remark 2.4.38.** (i) It is clear, that a similar discussion of the solutions \( g_z \) is possible, if \( a = -\infty \) or \( b = \infty \).

(ii) Note, that we can write \( g_z'(x) = \frac{(1-F(z)p(x)H(x)}{I(z)^2} \) for \( x \in (a, z) \) and \( g_z'(x) = \frac{-F(z)p(x)G(x)}{I(z)^2} \) for \( x \in (z, b) \), with the functions \( H \) and \( G \) from Lemma 2.4.18. For concrete distributions one may often prove, that \( g_z'(x) \) is increasing on \( (a, z) \) and decreasing on \( (z, b) \), but this seems to be hard to prove in generality, if it is true at all.

Finally, in our general setting, we will prove suitable “plug-in theorems” for exchangeable pairs satisfying our general regression property (2.12). As was observed in [Röl08] for the normal distribution, in case of univariate distributional approximations, one does not need the full strength of exchangeability, but equality in distribution of the random variables \( W \) and \( W' \) is sufficient. This may allow for a greater choice of admissible couplings in several situations, or at least, relaxes the verification of asserted properties.

In the following, let \( (\Omega, \mathcal{A}, P) \) be a probability space and let \( W, W' \) be real-valued random variables defined on this space such that \( W \overset{D}{=} W' \). Let, as before, \( \mu \) be our target distribution with support \( (a, b) \) fulfilling Condition 2.4.1. From now on we will assume, that the random variables \( W \) and \( W' \) only have values in an interval \( I \supseteq (a, b) \) where both functions \( \eta \) and \( \gamma \) are defined (recall that it might be the case that \( \eta \) can only be defined on \( (a, b) \)).

**Proposition 2.4.39.** Assume, that \( \gamma \) satisfies Condition 2.4.2 and that \( W \) is square integrable with \( E[|\gamma(W)|] < \infty \). Furthermore, let the pair \( (W, W') \) satisfy the general regression property (2.12). Let \( f : I \to \mathbb{R} \) be an absolutely continuous function with \( \|f\|_\infty, \|f'\|_\infty < \infty \), where \( f' \) is a given, Borel-measurable version of the derivative of \( f \). Then we have
2. Stein’s method for absolutely continuous univariate distributions

\[
\left| E[\eta(W)f'(W) + \gamma(W)f(W)] \right| \\
\leq \|f'\|_\infty E\left[ |\eta(W) - \frac{1}{2\lambda} E[(W' - W)^2]W] \right] \\
+ \frac{1}{\lambda} E\left[ (W' - W)^2 \int_0^1 (1 - s) |f'(W) - f'((W + s(W' - W))|ds \right] \\
+ \|f\|_\infty \lambda E[|R|]. 
\]

(2.42)

If \(f'\) is also absolutely continuous and \(\|f''\|_\infty < \infty\) for some Borel-measurable version \(f''\) of the second derivative, then we also have the bound

\[
\left| E[\eta(W)f'(W) + \gamma(W)f(W)] \right| \\
\leq \|f'\|_\infty E\left[ |\eta(W) - \frac{1}{2\lambda} E[(W' - W)^2]W] \right] \\
+ \|f''\|_\infty \frac{\lambda}{6\lambda} E[|W' - W|^3] \\
+ \|f\|_\infty \lambda E[|R|]. 
\]

(2.43)

Proof. Let \(x_0\) be as in Condition 2.4.2 and define the function \(G : I \to \mathbb{R}\) by \(G(x) := \int_{x_0}^x f(y)dy\). Then, by a suitable version of Taylor’s formula, for each \(x, x' \in I\) we have

\[
G(x') - G(x) = G'(x)(x' - x) + \int_x^{x'} (x' - t)G''(t)dt \\
= f(x)(x' - x) + \int_x^{x'} (x' - t)f'(t)dt \\
= f(x)(x' - x) + (x' - x)^2 \int_0^1 (1 - s)f'(x + s(x' - x))ds .
\]

Hence, by distributional equality, we obtain

60
2.4. Modification of Stein’s equation by exchangeable pairs

\[ 0 = E[G(W')] - E[G(W)] \]
\[ = E[f(W)(W' - W)] + E[(W' - W)^2 \int_0^1 (1 - s)f'((W + s(W' - W)))ds] \]
\[ = E[f(W)E[W' - W|W]] + E[(W' - W)^2 \int_0^1 (1 - s)f'((W + s(W' - W)))ds] \]
\[ = \lambda E[f(W)\gamma(W)] + E[f(W)R] \]
\[ + E[(W' - W)^2 \int_0^1 (1 - s)f'((W + s(W' - W)))ds], \]
yielding

\[ E[f(W)\gamma(W)] = -\frac{1}{\lambda} E[(W' - W)^2 \int_0^1 (1 - s)f'((W + s(W' - W)))ds] - \frac{1}{\lambda} E[f(W)R]. \tag{2.44} \]

This immediately implies the identity

\[ E[\eta(W)f'(W) + \gamma(W)f(W)] \]
\[ = E\left[f'(W)(\eta(W) - \frac{1}{2\lambda}(W' - W)^2)\right] \]
\[ + \frac{1}{\lambda} E\left[(W' - W)^2 \int_0^1 (1 - s)\left(f'(W) - f'((W + s(W' - W)))\right)ds\right] \]
\[ - \frac{1}{\lambda} E[f(W)R] \]
\[ = E\left[f'(W)\left(\eta(W) - \frac{1}{2\lambda} E[(W' - W)^2|W]\right)\right] \]
\[ + \frac{1}{\lambda} E\left[(W' - W)^2 \int_0^1 (1 - s)\left(f'(W) - f'((W + s(W' - W)))\right)ds\right] \]
\[ - \frac{1}{\lambda} E[f(W)R]. \tag{2.45} \]

From (2.45) and the assumptions on \( f \) the bound (2.42) now easily follows. To prove (2.43) it suffices to observe that

\[ \left|f'(W) - f'((W + s(W' - W)))\right| \leq \|f''\|_\infty s|W' - W| \]

and \( \int_0^1 s(1 - s)ds = \frac{1}{6}. \) \( \square \)
2. Stein’s method for absolutely continuous univariate distributions

Remark 2.4.40. (i) From the first term on the right hand side of (2.42) we see, that the bound can only be useful, if
\[ E[(W' - W)^2 | W] \approx 2\lambda \eta(W). \]
Similarly, the third term reveals, that, indeed, \( R \) should be of smaller order than \( \lambda \).

(ii) The proof shows, that Proposition 2.4.39 can easily be generalized to the situation, where there is a \( \sigma \)-algebra \( \mathcal{F} \) with \( \sigma(W) \subseteq \mathcal{F} \subseteq \mathcal{A} \) and the more general regression property
\[ E[W' - W | \mathcal{F}] = \lambda \gamma(W) + R \quad (2.46) \]
with some \( \mathcal{F} \)-measurable remainder term \( R \) is satisfied.

(iii) If \( \mathcal{H} \) is some class of test functions, such that there are finite, positive constants \( c_0, c_1 \) and \( c_2 \) with \( \| g_h \|_{\infty} \leq c_0, \| g'_h \|_{\infty} \leq c_1 \) and \( \| g''_h \|_{\infty} \leq c_2 \) for each \( h \in \mathcal{H} \), then (2.43) immediately yields a bound on the distance
\[ d_{\mathcal{H}}(\mu, \mathcal{L}(W)) = \sup_{h \in \mathcal{H}} \left| E[h(W)] - E[h(Z)] \right|. \]

(iv) Suppose we are given a real-valued random variable \( W \), whose distribution we cannot compute and which we want to approximate by some other distribution, which still has to be chosen. Suppose \( W' \) is defined on the same probability space as \( W \), has the same distribution as \( W \) and that the pair \( (W, W') \) satisfies the regression property (2.12) und additionally
\[ \frac{1}{2\lambda} E[(W' - W)^2 | W] = \eta(W) + S, \quad (2.47) \]
with the same positive constant \( \lambda \), a negligible term \( S \) and some given function \( \eta \), then Proposition 2.4.39 suggests approximating \( \mathcal{L}(W) \) by the distribution \( \mu \) with density \( p \) given by relation 2.30, where \( I(x_0) \) is simply the unique normalizing constant. This suggestion is in the same spirit as the one given in the paper [CS11], but is indeed more general and, hence, allows for wider applicability. Hence, the postulate that the “law of large numbers” from Remark 2.3.2 should be satisfied within the density approach, if the pair \( (W, W') \) satisfies (2.8), may be replaced by the demand that a pair satisfying 2.12 should fulfill (2.47).
2.5. Stein's method for Beta distributions

In the following we will specialize the abstract theory from Section 2.4 to the family of Beta distributions. This class of distributions is rather large including as special cases the arcsine distribution, the semi-circle law and the uniform distribution over an interval. Usually, the family of Beta distributions is defined as a family of distributions on \([0, 1]\) depending on two parameters \(a, b > 0\). Recall, that for \(a, b > 0\), the Beta function \(B(a, b) := \int_0^1 x^{a-1}(1-x)^{b-1} \, dx\) is well-defined and hence, the function \(q_{a,b}(x) := \frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}1_{(0,1)}(x)\) is a probability density function. The corresponding distribution \(\nu_{a,b}\) is called the Beta distribution on \([0, 1]\) corresponding to the parameters \(a\) and \(b\).

However, in some cases, including the semicircular law, it is more convenient to consider the corresponding distribution on \([-1, 1]\). Using the substitution rule, it is a matter of routine to check, that for \(\alpha, \beta > -1\) the function \(p_{\alpha,\beta}(x) := C(\alpha, \beta)(1-x)^\alpha(1+x)^\beta 1_{(-1,1)}(x)\) is also a probability density function, where \(C(\alpha, \beta) := \frac{1}{B(\alpha+1, \beta+1)2^{\alpha+\beta+1}} = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}\). Here, we have used the well-known relation \(B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\) between the Beta function and Euler’s Gamma function.

The probability distribution corresponding to the density \(p_{\alpha,\beta}\) will be denoted by \(\mu_{\alpha,\beta}\) and is called the Beta distribution on \([-1, 1]\) corresponding to the parameters \(\alpha\) and \(\beta\). We denote its distribution function by \(F_{\alpha,\beta}\). Thus, \(F_{\alpha,\beta}(x) = \int_{-\infty}^x p_{\alpha,\beta}(t) \, dt, x \in \mathbb{R}\). One can easily see that if \(X \sim \nu_{a,b}\) and if we define \(Y := 2X - 1\), then \(Y \sim \mu_{-1,a-1}\). Conversely, if \(Y \sim \mu_{\alpha,\beta}\) and if \(X := \frac{Y+1}{2}\), then \(X \sim \nu_{\beta+1,\alpha+1}\). As this transformation is merely a matter of translation and scaling it will only be necessary to develop the theory for one of these intervals. For definiteness we choose the interval \([-1, 1]\) but will later on transfer our theory to the corresponding distributions on \([0, 1]\).

From now on, fix \(\alpha, \beta > -1\). In Example 2.1.1 we have already computed a Stein operator \(L\) for \(\mu_{\alpha,\beta}\). This was given by

\[
Lg(x) = (1 - x^2)g'(x) - (\alpha + \beta + 2) x g(x) + (\beta - \alpha) g(x)
\]

for smooth enough functions \(g\), yielding the Stein identity

\[
E\left[(1 - Z^2)g'(Z) - [(\alpha + \beta + 2)Z + (\alpha - \beta)] g(Z)\right] = 0,
\]

where \(Z \sim \mu_{\alpha,\beta}\).

Next we want to show that the operator \(L\) characterizes the distribution \(\mu_{\alpha,\beta}\). To do so, we introduce the function \(I_{\alpha,\beta}(x) := C(\alpha, \beta)(1-x)^{\alpha+1}(1+x)^{\beta+1}1_{(-1,1)}(x) = (1-x^2)p_{\alpha,\beta}(x)\) and the class of functions \(\mathcal{K}_{\alpha,\beta}\) consisting of all continuous and
2. Stein’s method for absolutely continuous univariate distributions

piecewise continuously differentiable functions $g : \mathbb{R} \to \mathbb{R}$ vanishing at infinity with $\int_{\mathbb{R}} |g'(x)|I_{\alpha,\beta}(x)dx < \infty$. Our observations lead to the following proposition.

**Proposition 2.5.1** (Stein characterization for $\mu_{\alpha,\beta}$). A real-valued random variable $X$ is distributed according to $\mu_{\alpha,\beta}$ if and only if for all functions $g \in \mathcal{K}_{\alpha,\beta}$ the expected values $E[\sqrt{1-X^2}g'(X)]$ and $E[(\alpha + \beta + 2)Xg(X) + (\alpha - \beta)g(X)]$ exist and coincide.

**Proof.** First, let $\mathcal{L}(X) = \mu_{\alpha,\beta}$ and let $g \in \mathcal{K}_{\alpha,\beta}$. By the hypothesis and the transformation formula, we have

\[
\begin{align*}
\infty &> \int_{\mathbb{R}} |g'(x)|I_{\alpha,\beta}(x)dx = C(\alpha, \beta) \int_{-1}^{1} |g'(x)|(1-x)^{\alpha+1}(1+x)^{\beta+1}dx \\
&= C(\alpha, \beta) \int_{-1}^{1} (1-x^2)|g'(x)|(1-x)^{\alpha}(1+x)^{\beta}dx = \int_{\mathbb{R}} |(1-x^2)g'(x)|p_{\alpha,\beta}(x)dx \\
&= E[(1-X^2)g'(X)].
\end{align*}
\]

Hence the expected value $E[(1-X^2)g'(X)]$ exists. Since $g$ is continuous, it is bounded on $[-1, 1]$ and so the expected value $E[(\alpha + \beta + 2)Xg(X) + (\alpha - \beta)g(X)]$ exists, too. Again, by the transformation rule and since $g$ and $I_{\alpha,\beta}$ are absolutely continuous on $[-1, 1]$ we can use integration by parts and have

\[
\begin{align*}
E[(1-X^2)g'(X)] &= \int_{\mathbb{R}} (1-x^2)g'(x)p_{\alpha,\beta}(x)dx = \int_{-1}^{1} g'(x)I_{\alpha,\beta}(x)dx \\
&= gI_{\alpha,\beta}|_{-1}^{1} - \int_{-1}^{1} g(x)I'_{\alpha,\beta}(x)dx \\
&= 0 - C(\alpha, \beta) \int_{-1}^{1} g(x)[(\beta + 1)(1+x)^{\beta}(1-x)^{\alpha+1} - (\alpha + 1)(1-x)^{\alpha}(1+x)^{\beta+1}]dx \\
&= - \int_{-1}^{1} g(x)[\beta - \alpha - x(\beta + \alpha + 2)]C(\alpha, \beta)(1-x)^{\alpha}(1+x)^{\beta}dx \\
&= \int_{\mathbb{R}} g(x)[\alpha - \beta + x(\alpha + \beta + 2)]p_{\alpha,\beta}(x)dx \\
&= E[(\alpha + \beta + 2)Xg(X) + (\alpha - \beta)g(X)].
\end{align*}
\]

For the converse fix an arbitrary $z \in \mathbb{R} \setminus \{-1, +1\}$ and consider the solution $g_z$ to Stein’s equation

\[
(1-x^2)g'(x) - (\alpha + \beta + 2)xg(x) + (\beta - \alpha)g(x) = h_z(x) - \mu_{\alpha,\beta}((-\infty, z]), \quad (2.48)
\]
where \( h_z := 1_{(-\infty,z]} \). It will be shown in Proposition 2.5.2 below that \( g_z \in K_{\alpha,\beta} \), so that by hypothesis we have

\[
0 = E[(1 - X^2)g'_z(X) - (\alpha + \beta + 2)Xg_z(X) + (\beta - \alpha)g_z(X)]
= E[h_z(X) - \mu_{\alpha,\beta}((-\infty,z])] \\
= P(X \leq z) - \mu_{\alpha,\beta}((-\infty,z)).
\]

Since \( z \in \mathbb{R} \setminus \{-1, +1\} \) was arbitrary and by continuity from the right of distribution functions, the proof is complete. \( \square \)

Since we have fixed the parameters \( \alpha \) and \( \beta \), henceforth we may and will suppress them as sub-indices at objects which might well depend on them (for example we will simply write \( p \) for \( p_{\alpha,\beta} \) and so on). As we would like to use the theory from Section 2.4 we have to make sure, that our Stein identity for the Beta distribution fits into this framework, i.e. that relation (2.18) is satisfied with \( \eta(x) = 1 - x^2 \) and

\( \gamma(x) = -(\alpha + \beta + 2)x + (\beta - \alpha) = -[(\alpha + \beta + 2)x + (\alpha - \beta)] = (\alpha + \beta + 2)(E[Z] - x), \)

where we have used, that \( E[Z] = \frac{\beta - \alpha}{\alpha + \beta + 2} \). In principle, this is clear, because we have just established a Stein characterization for \( \mu \) and given the density \( p \) and the function \( \gamma \), the corresponding \( \eta \) is, of course, unique. However, we give a formal proof.

According to (2.18) we must show that

\[
(1 - x^2)p(x) = \int_{-1}^{x} \left[-(\alpha + \beta + 2)t + (\beta - \alpha)\right]p(t)dt \tag{2.49}
\]

holds for all \( x \in (-1, 1) \). First note, that

\[ \psi(x) = \frac{p'(x)}{p(x)} = \frac{-(\alpha + \beta)x + \beta - \alpha}{1 - x^2}. \]

Differentiating the left hand side of (2.49), we obtain

\[
dx{(1 - x^2)p(x)} = -2xp(x) + (1 - x^2)p'(x) = p(x)(-2x + (1 - x^2)\psi(x)) \\
= p(x)[-(\alpha + \beta + 2)x + (\beta - \alpha)]
\]

which is of course the derivative of the right hand side, too. Since

\[
\lim_{x \searrow -1} (1 - x^2)p(x) = 0 = \lim_{x \searrow -1} \int_{-1}^{x} \left[-(\alpha + \beta + 2)t + (\beta - \alpha)\right]p(t)dt,
\]

65
2. Stein’s method for absolutely continuous univariate distributions

relation (2.49) is proved. Note, that we may extend the functions $\gamma$ and $\eta$ to functions on $\mathbb{R}$ by just the same “analytical” expressions as above. Next we will show that all the conditions from Section 2.4 are satisfied in the special case of Beta distributions. It is easy to see, that Conditions 2.4.1, 2.4.2, 2.4.15 and 2.4.24 are satisfied by the functions $\gamma$, $\eta$ and $p$.

Condition 2.4.7 is also satisfied but need not be proved, because its most important conclusion, namely that $\eta(1) = \eta(-1) = 0$ is clear from the above discussion. To verify Conditions 2.4.25, 2.4.27 and 2.4.30, we must first define the functions $F_l$ on $(-\infty, -1)$ and $F_r$ on $(1, \infty)$. We claim, that the functions

$$F_l(x) := (\alpha + 1) \log(1 - x) + (\beta + 1) \log(-1 - x), \quad x < -1$$

and

$$F_r(x) := (\alpha + 1) \log(x - 1) + (\beta + 1) \log(1 + x), \quad x > 1$$

do the job. In fact,

$$F'_l(x) = \frac{\alpha + 1}{1 - x}(-1) + \frac{\beta + 1}{-1 - \alpha}(-1) = \frac{-(\alpha + \beta + 2)x + \beta - \alpha}{1 - x^2} = \frac{\gamma(x)}{\eta(x)}$$

for each $x < -1$ and similarly $F'_r(x) = \frac{\gamma(x)}{\eta(x)}$ for $x > 1$. From these two functions we obtain the functions $q_l : (-\infty, -1) \rightarrow \mathbb{R}$ and $q_r : (1, \infty) \rightarrow \mathbb{R}$ defined by

$$q_l(x) := \exp(F_l(x)) \frac{\eta(x)}{\eta(x)} = \frac{(1 - x)^{\alpha+1}(-1 - x)^{\beta+1}}{1 - x^2} = (1 - x)\alpha(-1 - x)\beta$$

and

$$q_r(x) := \exp(F_r(x)) \frac{\eta(x)}{\eta(x)} = \frac{(x - 1)^{\alpha+1}(1 + x)^{\beta+1}}{1 - x^2} = (x - 1)\alpha(1 + x)^\beta.$$

These two functions are of course locally integrable and hence, Condition 2.4.25 is satisfied. Since

$$\lim_{x \to -1} F_l(x) = -\infty = \lim_{x \to 1} F_r(x)$$

and

$$\lim_{x \to -\infty} F_l(x) = +\infty = \lim_{x \to \infty} F_r(x),$$

Conditions 2.4.27 and 2.4.30 also hold. Consequently, all results from Section 2.4 are valid in particular for the case of Beta distributions.
2.5. Stein’s method for Beta distributions

Now, let $h : \mathbb{R} \to \mathbb{R}$ be a given Borel-measurable test function with
\[
\int_{\mathbb{R}} |h(x)| p(x) \, dx < \infty, \quad \int_{x}^{-1} |h(x)| (-q_l(x)) \, dx < \infty \quad \text{for each } x < -1 \quad \text{and}
\]
\[
\int_{1}^{x} |h(x)| (-q_r(x)) \, dx < \infty \quad \text{for each } x > 1 \quad \text{and let } \tilde{h} = h - \mu(h). \quad \text{Then, according}
\]
to the theory from Section 2.4, the corresponding Stein solution is given by the
function $g_h : \mathbb{R} \to \mathbb{R}$ with

\[
g_h(x) = \begin{cases} 
\frac{\exp(-F_l(x)) \int_{-1}^{x} \tilde{h}(t) q_l(t) \, dt}{h(-1) - \mu(h)}, & x < -1 \\
\frac{1}{\eta(x) p(x)} \int_{-1}^{x} \tilde{h}(t) p(t) \, dt, & -1 < x < 1 \\
\frac{\exp(-F_r(x)) \int_{1}^{x} \tilde{h}(t) q_r(t) \, dt}{h(1) - \mu(h)}, & x = 1 \\
\end{cases}
\]

(2.50)

Here, the values of $g_h$ at ±1 are arbitrary, but they are chosen such that $g_h$
is continuous, whenever $h$ is continuous at ±1. This follows immediately from
Proposition 2.4.28.

The next result completes the proof of Proposition 2.5.1 by showing that the
solution $g_z$ is in the class $K_{a,\beta}$ whenever $z \neq \pm 1$.

**Proposition 2.5.2.** For each $z \in \mathbb{R} \setminus \{-1, +1\}$ the solution $g_z$ from Proposition
2.4.37 belongs to the class $K_{a,\beta}$, which was defined above before Proposition 2.5.1.

**Proof.** Let us focus on the case $z \in (-1, 1)$, the cases $z < -1$ and $z > 1$ being
similar and even easier. Of course, $g_z$ is continuously differentiable on each of
the intervals $(-\infty, -1)$, $(-1, z)$, $(z, 1)$ and $(1, \infty)$ and from Proposition 2.4.28 we
know that $g_z$ is continuous on $\mathbb{R}$ since $z \neq \pm 1$. By a result from calculus, to show
that, for example, $g_z$ is $C^1$ on $[-1, z]$, it is sufficient to prove that $\lim_{x \searrow -1} g_z'(x)$
exists in $\mathbb{R}$. To this end we recall from Remark 2.4.38 that for $-1 < x < z$ we have

\[
g_z'(x) = (1 - F(z)) \frac{p(x) H(x)}{I(x)^2} = (1 - F(z)) \frac{H(x)}{(1 - x^2)^2 p(x)}.
\]

Using de l’Hôpital’s rule A.1.5 and Lemma 2.5.5 (a) below, we obtain

\[
\lim_{x \searrow -1} g_z'(x) = (1 - F(z)) \lim_{x \searrow -1} \frac{H(x)}{(1 - x^2)^2 p(x)}
\]

\[
= (1 - F(z)) \lim_{x \searrow -1} \frac{(\alpha + \beta + 2) F(x)}{(1 - x^2) p(x) [\beta - \alpha - (\alpha + \beta + 4)x]}
\]

67
2. Stein’s method for absolutely continuous univariate distributions

\[
\begin{align*}
1 - F(z) - \frac{(\alpha + \beta + 2)F(x)}{2\beta + 4} & \Rightarrow \\
1 - F(z) - \frac{(\alpha + \beta + 2)p(x)}{2\beta + 4} & \Rightarrow \\
\frac{(1 - F(z))(\alpha + \beta + 2)}{(2\beta + 4)(2\beta + 2)} & .
\end{align*}
\]

Hence, \(g_z\) is continuously differentiable on \([-1, z]\). That \(g_z\) is continuously differentiable on the intervals \((-\infty, 1]\), \([z, 1]\) and \([1, \infty)\) can be proved similarly by using the representations for \(g_z'(x)\) from Proposition 2.4.37. It also turns out, that the right and left hand derivatives of \(g_z\) at \(\pm 1\) coincide, so that it is in fact differentiable at these points. So, we conclude, that \(g_z\) is continuous and piecewise continuously differentiable on \(\mathbb{R}\). Next, we show that \(g_z\) vanishes at infinity. This even applies to any bounded test function \(h\), since e.g. for \(x > 1\) we have

\[
|g_h(x)| = \left| \exp(-F_r(x)) \int_1^x \tilde{h}(t)q_r(t)dt \right| \leq \|	ilde{h}\|_\infty \int_1^x q_r(t)dt (x-1)^{\alpha+1}(1+x)^{\beta+1}
\]

and by de l’Hôpital’s rule

\[
\lim_{x \to \infty} \frac{\int_1^x q_r(t)dt}{(x-1)^{\alpha+1}(1+x)^{\beta+1}} = \lim_{x \to \infty} \frac{(x-1)^{\alpha+1}(1+x)^{\beta+1}}{(x-1)^{\alpha+1}(1+x)^{\beta+1}} = \lim_{x \to \infty} \frac{1}{(x-1)^{\alpha+1}(1+x)^{\beta+1}} = 0.
\]

Hence, \(\lim_{x \to \infty} g_h(x) = 0\). Similarly, one can prove that \(\lim_{x \to -\infty} g_h(x) = 0\).

It remains to show that

\[
\int_{-1}^{1} |g'_h(x)| \cdot I_{\alpha, \beta}(x)dx = \int_{-1}^{1} |g'_h(x)|(1 - x^2)p(x)dx < \infty .
\]

To this end, it suffices to see that the function \(|g'_h(x)|(1 - x^2)p(x)\) is bounded on \((-1, z)\) and on \((z, 1)\). Since it is continuous on \((-1, z]\) and on \((z, 1)\) (where \(g'_z(z) := \lim_{x \to z} g'_z(x)\) for definiteness), this claim will follow if we have proved that \(\lim_{x \to \pm 1} g'_z(x)(1 - x^2)p(x) = 0\). For \(x \in (-1, z)\) we have

\[68\]
2.5. Stein’s method for Beta distributions

\[ g'_z(x)(1 - x^2)p(x) = (1 - F(z)) \frac{H(x)}{1 - x^2} \quad \text{and} \]

\[ \lim_{x \searrow 1} \frac{H(x)}{1 - x^2} = \lim_{x \searrow 1} \frac{(\alpha + \beta + 2)F(x) - 2x}{1 - x^2} = 0, \]

proving the claim for \(-1\). Since it may be proved analogously for \(+1\) the proof is complete.

Next, we will derive some results for the solutions \(g_h\) from corresponding results in Section 2.4.

**Proposition 2.5.3.** Let \(h : \mathbb{R} \to \mathbb{R}\) be bounded and Borel-measurable and let \(m\) be a median for \(\mu\). Then, we have the bound:

\[ \|g_h\|_{\infty} \leq \|h - \mu(h)\|_{\infty} \max \left( \frac{1}{2(1 - m^2)p(m)}, \frac{1}{2\beta + 2}, \frac{1}{2\alpha + 2} \right) \]

**Proof.** Since \(\gamma(-1) = 2\beta + 2\) and \(\gamma(1) = -2\alpha - 2\), this immediately follows from Proposition 2.4.31.

**Proposition 2.5.4.** Let \(h : \mathbb{R} \to \mathbb{R}\) be Lipschitz-continuous. Then, we have the following bounds:

(a) \(\|g_h\|_{\infty} \leq \frac{\|h\|_{\infty}}{\alpha + \beta + 2}\)

(b) There exists a constant \(K_1\), only depending on \(\alpha\) and \(\beta\) such that \(\|g_h\|_{\infty} \leq K_1\|h\|_{\infty}\).

Before giving the proof, we will formulate a useful lemma, which actually fits better into Section 2.4.

**Lemma 2.5.5.** The functions \(p, q_l, q_r\) and \(\eta\) satisfy the following equations for each integer \(k \geq 1\):

(a) \(\frac{d}{dx}(\eta(x)^k p(x)) = \eta(x)^{k-1} p(x) [(k - 1)\eta'(x) + \gamma(x)] \) for each \(x \in (-1, 1)\)

(b) \(\frac{d}{dx}(\eta(x)^k q_l(x)) = \eta(x)^{k-1} q_l(x) [(k - 1)\eta'(x) + \gamma(x)] \) for each \(x \in (-\infty, -1)\)

(c) \(\frac{d}{dx}(\eta(x)^k q_r(x)) = \eta(x)^{k-1} q_r(x) [(k - 1)\eta'(x) + \gamma(x)] \) for each \(x \in (1, \infty)\)
2. Stein’s method for absolutely continuous univariate distributions

Proof. First we prove (a). By (2.17) we have, multiplying by \( p(x) \),
\[
\eta(x)p'(x) = p(x)(\gamma(x) - \eta'(x)).
\]

Hence, by the product rule we obtain
\[
\frac{d}{dx}(\eta(x)^k p(x)) = k\eta(x)^{k-1}\eta'(x)p(x) + \eta(x)^kp'(x)
\]
\[
= \eta(x)^{k-1}[k\eta'(x)p(x) + \eta(x)p'(x)]
\]
\[
= \eta(x)^{k-1}p(x)[k\eta'(x) + \gamma(x) - \eta'(x)]
\]
\[
= \eta(x)^{k-1}p(x)[(k-1)\eta'(x) + \gamma(x)],
\]
proving (a). As to (b), we observe that by the definition of \( q_l = \frac{\exp F_l}{\eta} \) we have on the one hand
\[
\frac{d}{dx}(\eta(x)q_l(x)) = \frac{d}{dx}\exp(F_l(x)) = \exp(F_l(x))\frac{\gamma(x)}{\eta(x)}
\]
and on the other hand, by the product rule,
\[
\frac{d}{dx}(\eta(x)q_l(x)) = \eta'(x)q_l(x) + q_l'(x)\eta(x).
\]
These two equations yield
\[
\eta(x)q_l'(x) = \gamma(x)q_l(x) - \eta'(x)q_l(x) = q_l(x)(\gamma(x) - \eta'(x)).
\]

Now the proof follows the lines of proof for \( p \) as above. Similarly one may prove (c).

Proof of Proposition 2.5.4. Assertion (a) immediately follows from Corollary 2.4.36 (a) since in this case \( c = \alpha + \beta + 2 \). Now, we turn to the proof of (b). First, consider \( x \in (-1, 1) \). By Corollary 2.4.21 (b) we have for \( x \in (-1, 1) \):
\[
|g_h'(x)| \leq \frac{2\|h'\|_{\infty}}{\alpha + \beta + 2} \frac{H(x)G(x)}{I(x)\eta(x)}.
\]

For \( x \in (-1, 1) \) let
\[
S(x) := \frac{H(x)G(x)}{I(x)\eta(x)} = \frac{H(x)G(x)}{\eta(x)^2 p(x)}.
\]

Page 70
Then $S$ is continuous on $(-1, 1)$ and hence, to show that $S$ is bounded on $(-1, 1)$, it suffices to prove that $S$ has finite limits at $\pm 1$. Since $\lim_{x \to -1} G(x) = \gamma(-1) = 2\beta + 2$ we obtain, using Lemma 2.5.5 and de l'Hôpital’s rule:

\[
\lim_{x \to -1} S(x) = \lim_{x \to -1} G(x) \frac{H(x)}{\eta(x)^2 p(x)} = \gamma(-1) \frac{(\alpha + \beta + 2)F(x)}{\eta(x)^2 p(x)}
\]

\[
= \frac{\gamma(-1)(\alpha + \beta + 2)}{\lim_{x \to -1}[\eta'(x) + \gamma(x)]} \lim_{x \to -1} \frac{F(x)}{\eta(x)p(x)}
\]

\[
= \frac{\gamma(-1)(\alpha + \beta + 2)}{2\beta + 4} \frac{1}{\lim_{x \to -1} \gamma(x)} = \frac{\gamma(-1)(\alpha + \beta + 2)}{2\beta + 4} \frac{1}{\gamma(-1)}
\]

\[
= \frac{\alpha + \beta + 2}{2\beta + 4} < \infty
\]

Here we have used, that $H'(x) = -\gamma'(x) F(x) = (\alpha + \beta + 2) F(x)$. Similarly one shows that

\[
\lim_{x \to 1} S(x) = \frac{\alpha + \beta + 2}{2\alpha + 4}.
\]

Thus, we have shown that $\sup_{x \in (-1, 1)} S(x) < \infty$. Now, we consider $x \in (-\infty, -1)$. From Corollary 2.4.36 (b) we have

\[
|g_h(x)| \leq 2\|h''\|_{\infty} \frac{\gamma(x)(-xQ_l(x) + \int_a^x t q_l(t) dt)}{-\eta(x)I_l(x)}.
\]

For $x \in (-\infty, -1)$ we consider the function

\[
S_l(x) := \frac{\gamma(x)(-xQ_l(x) + \int_a^x t q_l(t) dt)}{-\eta(x)I_l(x)} = \frac{\gamma(x) \int_a^x Q(t) dt}{\eta(x)^2 q_l(x)}.
\]

Clearly, $S_l$ is a continuous function on $(-\infty, -1)$. To show that it is bounded, it thus suffices to prove that $\lim_{x \to -1} S_l(x) < \infty$ and $\lim_{x \to -\infty} S_l(x) < \infty$. Using Lemma 2.5.5 and de l'Hôpital’s rule, we obtain
2. Stein’s method for absolutely continuous univariate distributions

\[
\lim_{x \nearrow -1} S_l(x) = \lim_{x \nearrow -1} \gamma(x) \lim_{x \nearrow -1} \int_a^x Q(t) dt \frac{Q_l(x)}{\eta(x) q_l(x)}
\]

\[
= \gamma(-1) \lim_{x \nearrow -1} \frac{\eta(x) q_l(x)[\eta'(x) + \gamma(x)]}{\eta(x) q_l(x)} \lim_{x \nearrow -1} \frac{Q_l(x)}{\eta(x) q_l(x)}
\]

\[
= \gamma(-1) \frac{2\beta + 4}{2\beta + 4} \lim_{x \nearrow -1} \frac{q_l(x)}{\gamma(x) q_l(x)} \frac{1}{\gamma(x)} = \frac{1}{2\beta + 4} < \infty .
\]

Next, we will show that \(\lim_{x \to -\infty} S_l(x) = 0\). We actually have

\[
\lim_{x \to -\infty} \frac{\gamma(x)}{\eta(x)} = \lim_{x \to -\infty} \frac{-(\alpha + \beta + 2)x + \beta - \alpha}{1 - x^2} = 0
\]

and by de l’Hôpital’s rule, also

\[
\lim_{x \to -\infty} \int_a^x Q(t) dt \frac{q_l(x)}{\eta(x) q_l(x)} = \lim_{x \to -\infty} \frac{q_l(x)}{q_l(x) \gamma(x)} = \lim_{x \to -\infty} \frac{1}{\gamma(x)} = 0 .
\]

Here we have used that \(\eta(x) q_l(x) = -(1 - x)^{\alpha + 1}(-1 - x)^{\beta + 1} \to -\infty\) as \(x \to -\infty\). Hence, \(\lim_{x \to -\infty} S_l(x) = 0\) and \(\sup_{x \in (-\infty, -1)} S_l(x) < \infty\). Since we can show in a similar manner that \(\sup_{x \in (1, \infty)} S_r(x) < \infty\), where \(S_r\) is defined in the obvious way, the proof is complete. \(\square\)

Now, consider a twice differentiable function \(h : \mathbb{R} \to \mathbb{R}\) with bounded first and second derivative. Recall the discussion following Remark 2.4.22. As was explained there, in the case of Beta distributions, we have \(\tilde{\mu}_{\alpha, \beta} = \mu_{\alpha + 1, \beta + 1}\) and we have to show that the function \(\tilde{g} := g'_h\) satisfies the Stein identity for \(\mu_{\alpha + 1, \beta + 1}\), i.e. for \(Y \sim \mu_{\alpha + 1, \beta + 1}\) we have

\[
E\left[ (1 - Y^2)\tilde{g}'(Y) + [-(\alpha + \beta + 4)Y + \beta - \alpha]\tilde{g}(Y) \right] = 0 . \quad (2.51)
\]

The following lemma will be useful.

**Lemma 2.5.6.** For the Beta distribution \(\mu_{\alpha, \beta}\) and a given bounded, Borel-measurable function \(u : \mathbb{R} \to \mathbb{R}\), the Stein equation

\[
\eta(x) f'(x) + \gamma(x) f(x) = u(x)
\]

has a bounded solution \(f\) on \((-1, 1)\) if and only if \(E[u(Z)] = 0\).
2.5. Stein’s method for Beta distributions

Proof. If $E[u(Z)] = 0$, then the usual Stein solution is bounded on $(-1, 1)$ by Proposition 2.5.3. For the converse, let us assume that $E[u(Z)] \neq 0$. As was already noted in Section 2.4, the solutions of the homogeneous equation corresponding to (2.52) are exactly the multiples of $\frac{1}{\eta p}$. Thus, every solution $f$ of (2.52) has the form

$$f(x) = \int_{-1}^{x} u(t)p(t)dt + c \eta(x)p(x), \quad x \in (-1, 1),$$

for some $c \in \mathbb{R}$. If $c = 0$, then

$$\lim_{x \nearrow 1} f(x) = \lim_{x \nearrow 1} \int_{-1}^{x} u(t)p(t)dt = \pm \infty,$$

since $E[u(Z)] \neq 0$. Hence, $f$ is unbounded near $1$. If $c \neq 0$ we have

$$\lim_{x \searrow -1} f(x) = \lim_{x \searrow -1} \frac{c}{\eta(x)p(x)} = \pm \infty.$$

So, $f$ is unbounded near $-1$. Hence, in any case $f$ is unbounded on $(-1, 1)$.

Now, from Proposition 2.5.4 we know that $\tilde{g}$ is bounded and from (2.31) we know that $\tilde{g}$ satisfies the Stein equation corresponding to $\tilde{\mu}$ for the test function $h_2(x) = h'(x) - \gamma'(x)g_b(x)$. By Lemma 2.5.6 we thus have that $\int_{\mathbb{R}} h_2(x)d\mu_{\alpha+1,\beta+1}(x) = 0$. Hence, $\tilde{g}$ must be the only bounded solution to the Stein equation for $\mu_{\alpha+1,\beta+1}$ corresponding to the test function $h_2$. Since

$$h_2(x) = h'(x) - \gamma'(x)g_b(x) = h'(x) + (\alpha + \beta + 2)g_b(x)$$

we have

$$h'_2(x) = h''(x) + (\alpha + \beta + 2)g'_b(x)$$

and from Proposition 2.5.4 we see that $h_2$ is Lipschitz with minimal Lipschitz constant

$$\|h'_2\|_\infty \leq \|h''\|_\infty + (\alpha + \beta + 2)K_1(\alpha, \beta)\|h'\|_\infty,$$

where the constant $K_1(\alpha, \beta)$ from Proposition 2.5.4 only depends on $\alpha$ and $\beta$. Applying Proposition 2.5.4 again, this time to the distribution $\mu_{\alpha+1,\beta+1}$ and the Stein solution $\tilde{g}$, we obtain

$$\|g''_h\|_\infty = \|\tilde{g}'\|_\infty \leq K_1(\alpha + 1, \beta + 1)\|h'_2\|_\infty \leq K_1(\alpha + 1, \beta + 1)(\|h''\|_\infty + (\alpha + \beta + 2)K_1(\alpha, \beta)\|h'\|_\infty).$$
2. Stein’s method for absolutely continuous univariate distributions

Hence, there is a constant $K_2$ depending only on $\alpha$ and $\beta$ such that
\[ \|g''_h\|_{\infty} \leq K_2(\|h'||_{\infty} + \|h''|_{\infty}) \]
for all twice differentiable functions $h$ with bounded first and second derivative. We have thus proved the following proposition.

**Proposition 2.5.7.** There exists a finite constant $K_2$ depending only on $\alpha$ and $\beta$ such that for each twice differentiable function $h : \mathbb{R} \rightarrow \mathbb{R}$ with bounded first and second derivative we have the bound
\[ \|g''_h\|_{\infty} \leq K_2(\|h'||_{\infty} + \|h''|_{\infty}) \cdot \]

Now we are in the position to provide a “plug-in theorem” for the Beta approximation using exchangeable pairs.

**Theorem 2.5.8.** Let $W, W'$ be identically distributed, real-valued random variables on a common probability space $(\Omega, \mathcal{A}, P)$ satisfying the regression property
\[ E[W' - W|W] = \lambda \gamma(W) + R \]
for some constant $\lambda > 0$ and a random variable $R$. Then for each twice differentiable function $h$ with bounded first and second derivative and with $E[|h(W)|] < \infty$ we have the bound
\[
|E[h(W)] - \mu_{\alpha, \beta}(h)| \\
\leq K_1 \|h'||_{\infty} E[\left| (1 - W^2) - \frac{1}{2\lambda} E[(W' - W)^2|W] \right|] \\
+ \frac{K_2(\|h'||_{\infty} + \|h''|_{\infty})}{6\lambda} E[|W' - W|^3] \\
+ \frac{\|h'||_{\infty}}{(\alpha + \beta + 2)\lambda} E[|R|] ,
\]
where the constants $K_1$ and $K_2$ are from Propositions 2.5.4 and 2.5.7, respectively.

**Proof.** This immediately follows from Propositions 2.4.39, 2.5.4, 2.5.7 and since $g_h$ is a solution to Stein’s equation (2.14).

In the following we will transfer the developed theory to the Beta distributions $\nu_{a,b}$ on $[0,1]$. We start with the Stein identity for $\nu_{a,b}$, where $a, b > 0$ are fixed parameters. Let $X \sim \nu_{a,b}$, then $Y := 2X - 1 \sim \mu_{b-1, a-1}$ and hence for each smooth enough function $f$ we have
2.5. Stein’s method for Beta distributions

\[ 0 = E[(1 - Y^2)f'(Y) - (a + b)Yf(Y) + (a - b)f(Y)]. \]

Let \( \hat{f}(x) := f(2x - 1) \). Then \( \hat{f}'(x) = 2f'(2x - 1) \) and \( \hat{f}(X) = f(Y) \). Hence, we obtain

\[
0 = E[(1 - Y^2)f'(Y) - (a + b)Yf(Y) + (a - b)f(Y)]
= E[4X(1 - X)f'(2X - 1) - (a + b)(2X - 1)f(2X - 1) + (a - b)f(2X - 1)]
= E[2X(1 - X)f'(X) - 2(a + b)X\hat{f}(X) + 2a\hat{f}(X)]
= 2E[X(1 - X)f'(X) - (a + b)X\hat{f}(X) + a\hat{f}(X)].
\]

So, a Stein identity for \( X \sim \nu_{a,b} \) is given by

\[
E\left[X(1 - X)f'(X) + \left[-(a + b)X + a\right]f(X)\right] = 0
\]

for all smooth enough functions \( f : \mathbb{R} \to \mathbb{R} \). Hence, for \( \nu_{a,b} \) the functions \( \eta \) and \( \gamma \) are given by \( \eta(x) = x(1 - x) \) and \( \gamma(x) = -(a + b)x + a = (a + b)(\frac{a}{a+b} - x) \).

Note that \( E[X] = \frac{a}{a+b} \) if \( X \sim \nu_{a,b} \). Having derived the Stein identity for \( \nu_{a,b} \), the Stein equation corresponding to a Borel-measurable test function \( h : \mathbb{R} \to \mathbb{R} \) with \( \int_{\mathbb{R}}|h(x)|\,d\nu_{a,b}(x) < \infty \) is given by

\[
x(1 - x)f'(x) + \left[-(a + b)x + a\right]f(x) = h(x) - \nu_{a,b}(h) =: \hat{h}(x). \tag{2.53}
\]

Let a test function \( h : \mathbb{R} \to \mathbb{R} \) with \( \int_{\mathbb{R}}|h(x)|\,d\nu_{a,b}(x) < \infty \) be given and let \( h_1(y) := h(\frac{y+1}{2}) \). Consider the above constructed solution \( g \) to the Stein equation for \( \mu_{b-1,a-1} \) corresponding to the test function \( h_1 \). In the following let the real variables \( x \) and \( y \) be related by \( y := 2x - 1 \) or \( x := \frac{y + 1}{2} \). Letting \( f(x) := 2g(2x - 1) \) and noting \( f'(x) = 4g(y) \) we obtain

\[
x(1 - x)f'(x) + \left[-(a + b)x + a\right]f(x)
= \frac{y+1}{2}(1 - \frac{y+1}{2})4g(y) + \left[-(a + b)(y + 1) + 2a\right]g(y)
= (1 - y^2)g(y) + \left[-(a + b)y + a - b\right]g(y)
= h_1(y) - \mu_{b-1,a-1}(h_1) = h(x) - \nu_{a,b}(h),
\]

since for any admissible function \( u : \mathbb{R} \to \mathbb{R} \) we have \( E[u(X)] = E[u_1(Y)] \) where \( u_1(y) := u(\frac{y+1}{2}) \). Hence, \( f \) is a solution to (2.53). Thus, we immediately get bounds on the solutions of the Stein equation for \( \nu_{a,b} \) from our above developed theory. In the following, we will always denote by \( f_h \) the Stein solution to (2.53) which is constructed in the explained way.
2. Stein’s method for absolutely continuous univariate distributions

**Proposition 2.5.9.** Let \( h : \mathbb{R} \to \mathbb{R} \) be Borel-measurable with \( \int_{\mathbb{R}} |h(x)| \, dv_{a,b}(x) < \infty \).

(a) If \( h \) is bounded, then \( \|f_h\|_{\infty} \leq \|h - \nu_{a,b}(h)\|_{\infty} \max\left(\frac{1}{2m(1-m)d_{a,b}(m)}, \frac{1}{a}, \frac{1}{b}\right) \), where \( m \) is a median for \( \nu_{a,b} \).

(b) If \( h \) is Lipschitz, then \( \|f_h\|_{\infty} \leq \frac{2}{a+b} \|h'\|_{\infty} \) and \( \|f'_h\|_{\infty} \leq C_1 \|h'\|_{\infty} \), where \( C_1 \) only depends on \( a \) and \( b \).

(c) If \( h \) is twice differentiable with bounded first and second derivative, then \( \|f''_h\|_{\infty} \leq C_2 \left( \|h''\|_{\infty} + \|h''\|_{\infty} \right) \), where \( C_2 \) only depends on \( a \) and \( b \).

**Proof.** Claim (a) follows from Proposition 2.4.31. Since \( f_h(x) = 2g_h(2x - 1) \) for all \( x \in \mathbb{R} \), (b) and (c) follow from Propositions 2.5.4 and 2.5.7.

Now, let \( V, V' \) be identically distributed, real-valued random variables on a common probability space \((\Omega, \mathcal{A}, P)\). For the approximation of \( \mathcal{L}(V) \) by \( \nu_{a,b} \) the general regression property from Section 2.4 is

\[
E[V' - V|V] = \lambda (a + b) \left( \frac{a}{a + b} - V \right) + R, \tag{2.54}
\]

where, again, \( \lambda > 0 \) is constant and \( R \) is a hopefully small remainder term. For the distribution \( \nu_{a,b} \) Theorem 2.5.8 becomes the following:

**Theorem 2.5.10.** Let \( V, V' \) be identically distributed, real-valued random variables on a common probability space \((\Omega, \mathcal{A}, P)\) satisfying equation (2.54). Then, for each twice differentiable function \( h : \mathbb{R} \to \mathbb{R} \) with bounded first and second derivative and with \( E[|h(V)|] < \infty \) we have the bound

\[
|E[h(V)] - \nu_{a,b}(h)| \\
\leq C_1 \|h\|_{\infty} E\left[|V(1 - V) - \frac{1}{2\lambda} E[(V' - V)^2|V]| \right] \\
+ \frac{C_2 (\|h'\|_{\infty} + \|h''\|_{\infty})}{6\lambda} E[|V' - V|^3] \\
+ \frac{2\|h'\|_{\infty}}{(a + b)\lambda} E[|R|],
\]

where the constants \( C_1 \) and \( C_2 \) are those from Proposition 2.5.9.

**Proof.** The assertion is clear from Propositions 2.4.39 and 2.5.9 and since \( f_h \) is a solution to Stein’s equation (2.53). \( \square \)
2.6. The Pólya urn model

In this section we will prove quantitative versions of the fact that the relative number of drawn red balls in a Pólya urn model converges in distribution to a suitable Beta distribution, if the number of total drawings tends to infinity. This model will serve as an application of our Stein method for the Beta distribution, as developed in Section 2.5. We start by introducing the stochastic model:

Imagine an urn containing at the beginning $r$ red and $w$ white balls and fix an integer $c > 0$. At each time point $n \in \mathbb{N}$ a ball is drawn at random from the urn, its color is noticed and this ball together with $c$ further balls of the same color is replaced to the urn. For each $n \in \mathbb{N}$ let $X_n$ be the indicator random variable of the event that the $n$-th drawn ball is a red one. Then $S_n := \sum_{j=1}^{n} X_j$ denotes the total number of drawn red balls among the first $n$ drawings. It is a well-known fact from elementary probability theory that for each $n \in \mathbb{N}$ and all $x_1, \ldots, x_n \in \{0, 1\}$ we have

$$P(X_1 = x_1, \ldots, X_n = x_n) = \frac{\prod_{i=0}^{k-1} (r + ci) \prod_{j=0}^{n-k-1} (w + cj)}{\prod_{l=0}^{n-1} (r + w + cl)},$$

where $k := \sum_{j=1}^{n} x_j$. This shows particularly that the sequence $(X_j)_{j \in \mathbb{N}}$ is exchangeable, meaning that its distribution is invariant under any permutation of finitely many indices.

It now follows, that for each $k = 0, \ldots, n$ we have

$$P(S_n = k) = \binom{n}{k} \frac{\prod_{i=0}^{k-1} (r + ci) \prod_{j=0}^{n-k-1} (w + cj)}{\prod_{l=0}^{n-1} (r + w + cl)},$$

or, with $a := \frac{r}{c}$ and $b := \frac{w}{c}$,

$$P(S_n = k) = \frac{\binom{n}{k}}{\binom{n-k}{a}} \cdot \frac{\binom{n-k}{b}}{\binom{n-k}{a-b}}.$$

The distribution of $S_n$ is usually referred to as the Pólya distribution with parameters $n$, $a$ and $b$. It is a well-known fact that the distribution of $\frac{1}{n} S_n$ converges weakly to the distribution $\nu_{a,b}$ as $n$ goes to infinity, where the Beta distribution $\nu_{a,b}$ was defined in Section 2.5. A convenient way to prove this weak convergence result is to use the formula

$$P(S_n = k) = \int_{0}^{1} b(k; n, p) \, d\nu_{a,b}(p), \quad (2.55)$$

77
2. Stein’s method for absolutely continuous univariate distributions

together with the weak law of large numbers for Bernoulli random variables to
deal with the binomial probabilities \( b(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k} \).

Formula (2.55) can be proved by a straight-forward computation using the relations \( B(a + 1, b) = \frac{a}{a + b} B(a, b) \) and \( B(a, b) = B(b, a) \) for the Beta function, where \( a, b > 0 \), and can also be viewed as a consequence of a special instance of de Finetti’s representation theorem for infinite exchangeable sequences. Note, however, that one generally does not know the corresponding mixing measure from de Finetti’s theorem and hence, identity (2.55) is not a direct consequence of this theorem.

From now on, we will present a Stein’s method proof of the above distributional convergence result and, as usual, also derive a rate of convergence. We will usually suppress the time index \( n \) and let \( V := V_n := \frac{1}{n} S_n \) denote the random variable of interest. For the construction of the exchangeable pair, we use the Gibbs sampling procedure (see Appendix A.2) with the slight simplification, that due to the exchangeability of \( X_1, \ldots, X_n \) we need not choose at random the index of the summand from \( S_n \), which has to be replaced. Instead, we will always replace \( X_n \) by \( X_n' \), which is constructed as follows:

Observe \( X_1 = x_1, \ldots, X_n = x_n \) and construct \( X_n' \) according to the distribution \( \mathcal{L}(X_n' | X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) \). Then, letting \( V' := V'_n := V - \frac{1}{n} X_n + \frac{1}{n} X_n' \), the pair \( (V, V') \) is exchangeable. In order to use Stein’s method of exchangeable pairs, we need to establish a suitable regression property. This is the content of the following proposition.

**Proposition 2.6.1.** The exchangeable pair \( (V, V') \) satisfies the regression property

\[
E[V' - V | V] = \frac{a + b}{n(a + b + n - 1)} \left( \frac{a}{a + b} - V \right) = \lambda \gamma_{a, b}(V),
\]

where \( \gamma_{a, b}(x) = (a + b)(\frac{a}{a + b} - x) \) and \( \lambda = \lambda_n = \frac{1}{n(a + b + n - 1)} \).

**Proof.** We have \( V' - V = \frac{X_n}{n} - \frac{X_n'}{n} \) and by exchangeability of \( X_1, \ldots, X_n \) it clearly holds that \( E[X_n | S_n] = E[X_j | S_n] \) for each \( 1 \leq j \leq n \). From \( S_n = E[S_n | S_n] = \sum_{j=1}^n E[X_j | S_n] = nE[X_n | S_n] \) it then follows that \( E[X_n | V] = E[X_n | S_n] = \frac{1}{n} S_n = V \). Also, by the definition of \( X_n' \) and since \( X_n' \) only assumes the values 0 and 1 we have for any \( x_1, \ldots, x_{n-1} \in \{0, 1\} \)

\[
E[X_n' | X_1 = x_1, \ldots, X_n = x_n] = E[X_n | X_1 = x_1, \ldots, X_{n-1} = x_{n-1}]
\]

\[
= P(X_n = 1 | X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \frac{r + c \sum_{j=1}^{n-1} x_j}{r + w + c(n - 1)}.
\]
and hence,

\[ E[X'_n|X_1, \ldots, X_n] = \frac{r + c \sum_{j=1}^{n-1} X_j}{r + w + c(n-1)} = \frac{r + cnV - cX_n}{r + w + c(n-1)}. \]

Thus, since \( \sigma(V) \subseteq \sigma(X_1, \ldots, X_n) \), we obtain

\[
E[X'_n|V] = E\left[E[X'_n|X_1, \ldots, X_n] | V\right] = \frac{r + c(n-1)V}{r + w + c(n-1)} = \frac{a + (n-1)V}{a + b + n - 1}.
\]

Finally, we have

\[
E[V' - V|V] = \frac{1}{n} E[X'_n - X_n|V] = \frac{1}{n} \frac{a + (n-1)V}{a + b + n - 1} - \frac{1}{n} V = \frac{a - (a + b)V}{n(a + b + n - 1)} = \frac{a + b}{n(a + b + n - 1)} \left( \frac{a}{a + b} - V \right),
\]

as was to be shown. \( \square \)

Next, we will compute the quantity \( E[(V' - V)^2|V] \).

**Proposition 2.6.2.** We have for the above constructed exchangeable pair \( (V, V') \)

\[
E[(V' - V)^2|V] = \frac{1}{n^2(a + b + n - 1)} \left( (2n + b - a)V - 2nV^2 + a \right)
\]

and hence

\[
\frac{1}{2\lambda} E[(V' - V)^2|V] = V(1 - V) + \frac{b - a}{2n} V + \frac{a}{2n}.
\]

**Proof.** From the general theory of Gibbs sampling (see Proposition A.2.3 (c)) we know, that

\[
E[(V' - V)^2|V] = \frac{1}{n^2} \left( E[X_n^2|V] + E[E[X'_n|X_1, \ldots, X_n]|V] - 2E[X_nE[X_n|X_1, \ldots, X_n]|V]\right).
\]

Since \( X_n^2 = X_n \) we have from the proof of Proposition 2.6.1 that
2. Stein’s method for absolutely continuous univariate distributions

\[ E[X_n^2|X_1,\ldots,X_{n-1}] = E[X_n|X_1,\ldots,X_{n-1}] = \frac{a + nV - X_n}{a + b + n - 1}, \]

and hence

\[ E[E[X_n^2|X_1,\ldots,X_{n-1}]|V] = \frac{a + (n-1)V}{a + b + n - 1}. \]

We also already know that \( E[X_n^2|V] = E[X_n|V] = V \). Finally, we compute

\[ E[X_nE[X_n|X_1,\ldots,X_{n-1}]|V] = E\left[\frac{aX_n + nVX_n - X_n^2}{a + b + n - 1} \bigg| V\right] \]
\[ = \frac{aV + nV^2 - V}{a + b + n - 1} = \frac{(a - 1)V + nV^2}{a + b + n - 1}. \]

Putting pieces together, we eventually obtain

\[ E[(V' - V)^2|V] = \frac{1}{n^2} \left(V + \frac{a + (n-1)V}{a + b + n - 1} - 2\frac{(a - 1)V + nV^2}{a + b + n - 1}\right) \]
\[ = \frac{1}{n^2(a + b + n - 1)} \left((2n + b - a)V - 2nV^2 + a\right). \]

The last assertion easily follows from this and from \( \lambda = \frac{1}{n(a+b+n-1)}. \quad \square \)

Recall that for the distribution \( \nu_{a,b} \) we have \( \eta(x) := \eta_{a,b}(x) = x(1-x) \) and hence, we obtain from Proposition 2.6.2 that

\[ E\left[|\eta(V) - \frac{1}{2\lambda} E[(V' - V)^2|V]|\right] = E\left[\frac{|a-b|}{2n} V - \frac{a}{2n}\right] \leq \frac{|a-b| + a}{2n}, \quad (2.57) \]

since \( |V| \leq 1 \). Similarly, since \( |V' - V| = \frac{1}{n}|X'_n - X_n| \leq \frac{1}{n} \) we have

\[ \frac{1}{6\lambda} E[|V' - V|^3] \leq \frac{n(a + b + n - 1)}{6} \frac{1}{n^3} = \frac{a + b + n - 1}{6n^2} \]
\[ = \frac{1}{6n} + \frac{a + b - 1}{6n^2} = O\left(\frac{1}{n}\right). \quad (2.58) \]

From Theorem 2.5.10 we can now conclude the following result.
2.6. The Pólya urn model

**Theorem 2.6.3.** For each twice differentiable function $h : \mathbb{R} \to \mathbb{R}$ with bounded first and second derivative we have

$$
\left| E[h(V)] - \nu_{a,b}(h) \right| \\
\leq \left( C_1 \|h'\|_{\infty} \left| \frac{a-b}{2} + a \right| + C_2 (\|h'\|_{\infty} + \|h''\|_{\infty}) \left( \frac{1}{6} + \frac{a + b - 1}{6n} \right) \right) \frac{1}{n}
$$

$$
= O\left( \frac{1}{n} \right),
$$

with the constants $C_1$ and $C_2$ from Proposition 2.5.9.

**Proof.** Since $V$ assumes only values in $[0, 1]$, the condition $E[|h(V)|] < \infty$ from Theorem 2.5.10 is trivially met. The assertion now follows immediately from Theorem 2.5.10, (2.57) and (2.58).

**Remark 2.6.4.** The result of this section appeared as an application of Stein's method for Beta distributions in the preprint [Döb12b]. A few days after [Döb12b] was on the arXiv, Goldstein and Reinert posted the preprint [GR12], in which they also prove rates of convergence of order $n^{-1}$ for the Pólya urn model by Stein's method, but even for the Wasserstein distance. Since they only bound the Stein solutions $f_h$ in the case that $a, b \geq 1$, they are able to compute a concrete constant of convergence. However, because of the restrictions $a, b \geq 1$ they only treat the case that after each drawing only one further ball of the same color is added to the urn. Using their technique of comparing with the Stein characterization of the discrete approximating distribution (see Section 2.8) together with the bounds from Proposition 2.5.9 (b) one can easily prove a rate of convergence of order $n^{-1}$ in the Wasserstein distance for arbitrary $c$, albeit without giving a concrete constant of convergence.
2. Stein’s method for absolutely continuous univariate distributions

2.7. Wigner’s semicircle law

In this section we use the results from Section 2.5 to derive a rate of convergence for Wigner’s famous semicircle law in the case of the GUE, the Gaussian Unitary Ensemble of random matrices. To be concrete, for \( n \in \mathbb{N} \) let \( X_{j,k}, Y_{j,k} \sim N(0, \frac{1}{4n}) \) \((1 \leq j < k \leq n)\) and \( X_{j,j} \sim N(0, \frac{1}{2n}) \) \((1 \leq j \leq n)\) be independent random variables. Then one says that the Hermitian matrix \( A := (A_{j,k})_{1 \leq j,k \leq n} \) with

\[
A_{j,k} := \begin{cases} 
X_{j,k} + iY_{j,k}, & 1 \leq j < k \leq n \\
X_{k,j} - iY_{k,j}, & 1 \leq k < j \leq n \\
X_{j,j}, & 1 \leq j = k \leq n
\end{cases}
\]

belongs to the GUE. Here and in what follows \( i := \sqrt{-1} \). Thus, \( A \) is a random Hermitian matrix of size \( n \times n \) and hence \( A \) has \( n \) real eigenvalues (with multiplicities) \( \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \), where each eigenvalue appears as often as its multiplicity indicates. Of course, the eigenvalues are also random and it is known that they are in fact measurable functions and hence real-valued random variables. A proof of this claim can for example be found in the introduction to random matrices [AGZ10]. Given the eigenvalues of \( A \) one defines their empirical distribution \( L_n \), which is a random probability measure on \( (\mathbb{R}, \mathcal{B}) \), by

\[
L_n := \frac{1}{n} \sum_{j=1}^n \delta_{\lambda_j}.
\] (2.59)

By the monotone convergence theorem one can easily prove, that also the mapping

\[
\mathcal{B} \ni B \mapsto \bar{L}_n(B) := E[L_n(B)] \in \mathbb{R}
\] (2.60)

is a probability measure on \( \mathbb{R} \), which is called the expected spectral distribution of \( A \). It was first proved in [Wig55] by Eugene Wigner via the method of moments that for real symmetric matrices with Bernoulli distributed entries, \( \bar{L}_n \) converges weakly to the semicircle distribution on \([-1, 1]\) as \( n \) goes to infinity. Later in [Wig58] Wigner noted that this result even holds for a quite general class of absolutely continuous distributions of the entries, containing the centered normal distributions. Nowadays it is known that weak convergence of \( \bar{L}_n \) to the semicircle distribution holds for quite arbitrary distributions of the entries, as long as the entries \( A_{j,k}, 1 \leq j \leq k \leq n \), are independent, \( A_{j,k}, 1 \leq j < k \leq n \), are identically distributed and certain moment conditions are satisfied. This is one instance of the so-called universality principle prevailing in random matrix theory. These Hermitian matrices are sometimes called Wigner random matrices.

Further research has been done to quantify the convergence to the semicircle law.
2.7. Wigner’s semicircle law

The aim of this section is to derive a rate of convergence result for Wigner’s semicircle law and the GUE by Stein’s method. In the GUE case the measure $\bar{L}_n$ is absolutely continuous with respect to Borel-Lebesgue measure with density $f_n : \mathbb{R} \rightarrow \mathbb{R}$ given by

$$f_n(x) := \frac{2}{\sqrt{n}} \sum_{j=0}^{n-1} P_j(2\sqrt{n}x)^2 \varphi(2\sqrt{n}x), \quad (2.61)$$

where $P_j$ is the $j$-th Hermite polynomial, orthonormalized with respect to the standard normal distribution $N(0,1)$ which has density

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

In [GT05] the authors derive the following third order ordinary differential equation for $f_n$:

$$0 = \frac{1}{16n^2} f''''(x) + (1 - x^2)f''(x) + xf'(x) =: (Af_n)(x)$$

As explained in Section 2.1 one may use integration by parts to compute the conjugate operator $A^*$ or $L := -A^*$, which is given by

$$(Lg)(x) := \frac{1}{16n^2} g''''(x) + (1 - x^2)g''(x) - 3xg(x)$$

and which serves as a Stein operator for $\bar{L}_n$. By the representation (2.61) of the density $f_n$ one easily sees, that the edge terms from the integration by parts formula vanish whenever $g$ belongs to the class $\mathcal{F}_n$ of all three times differentiable functions $g$ such that

$$\lim_{|x| \to \infty} x^k g^{(j)}(x) \exp(-2nx^2) = 0$$

for all $j \in \{0, 1, 2, 3\}$ and each integer $k \geq 0$. This observation immediately leads to the following Stein characterization for $\bar{L}_n$, which was also given in [GT05].

**Proposition 2.7.1** (Stein characterization for $\bar{L}_n$). A real-valued random variable $X$ has distribution $\bar{L}_n$ if and only if for all functions $g \in \mathcal{F}_n$

$$E \left[ \frac{1}{16n^2} g''''(X) + (1 - X^2)g''(X) - 3Xg(X) \right] = 0. \quad (2.62)$$
2. Stein’s method for absolutely continuous univariate distributions

The Stein identity (2.62) for \( \bar{L}_n \) thus differs from that of the semicircle law (see, e.g. Example 2.1.1 (b)) only by the term \( \frac{1}{16n^2}E[g''(X)] \). We may use this comparison of the two Stein identities to bound the distance between the distribution \( \bar{L}_n \) and the semicircle distribution. To do so, we must first generalize Proposition 2.5.7 to higher order derivatives.

**Proposition 2.7.2.** Consider the Beta distribution \( \mu := \mu_{\alpha,\beta} \), where \( \alpha, \beta > -1 \) are fixed and let \( m \geq 1 \). There exists a finite positive constant \( K_m \) such that for each \( m \)-times differentiable test function \( h : \mathbb{R} \to \mathbb{R} \) with \( \|h^{(j)}\|_{\infty} < \infty \) for \( j = 1, \ldots, m \) we have

\[
\|g_h^{(m)}\|_{\infty} \leq K_m \left( \sum_{j=1}^{m} \|h^{(j)}\|_{\infty} \right),
\]

where \( g_h \) is given by (2.50).

**Proof.** This follows from induction on \( m \) by the same argument that was used for the proof of Proposition 2.5.7 in the case \( m = 2 \). \( \square \)

Note that, for \( m \geq 3 \), an \( m \)-times differentiable function \( g \) with \( \|g^{(m)}\|_{\infty} < \infty \) automatically belongs to the class \( \mathcal{F}_n \) defined above. Hence, if \( h \) is three times differentiable with bounded first three derivatives \( h', h'' \) and \( h''' \), then, using (2.62), we have for \( X_n \sim \bar{L}_n \) and the semicircle distribution \( \mu_{\frac{1}{2}, \frac{1}{2}} \) on \([-1, 1]\) that

\[
\left| E[h(X_n)] - \mu_{\frac{1}{2}, \frac{1}{2}}(h) \right| = \left| E[(1 - X_n)^2 g_h'(X_n) - 3X_n g_h(X_n)] \right|
\]

\[
= \frac{1}{16n^2} \left| E[g_h'''(X_n)] \right| \leq \frac{\|g_h'''\|_{\infty}}{16n^2}
\]

\[
\leq \frac{K_3(\|h'\|_{\infty} + \|h''\|_{\infty} + \|h'''\|_{\infty})}{16n^2},
\]

with the constant \( K_3 \) from Proposition 2.7.2. Thus we have proved the following result.

**Theorem 2.7.3.** Let \( X_n \) have distribution \( \bar{L}_n \) and let \( h : \mathbb{R} \to \mathbb{R} \) have three bounded derivatives. Then we have

\[
\left| E[h(X_n)] - \mu_{\frac{1}{2}, \frac{1}{2}}(h) \right| \leq \frac{K(\|h'\|_{\infty} + \|h''\|_{\infty} + \|h'''\|_{\infty})}{16n^2}
\]

where \( K > 0 \) is a finite constant which does not depend on \( h \).

**Remark 2.7.4.** In [GT05] the authors show, that the Kolmogorov distance between \( \bar{L}_n \) and the semicircle distribution is of order \( n^{-1} \) and claim that this rate is optimal. Theorem 2.7.3 thus gives a better rate of convergence for a weaker metric.
In the remainder of this section we show how one can use mollifiers to obtain a rate of convergence of order \( n^{-2/3} \) in the Wasserstein distance between \( \bar{L}_n \) and \( \mu_{1,1} \).

Of course, the technique of convolving with a suitable mollifier is a well-known tool in classical analysis and particularly in the theory of partial differential equations (see, e.g., [Fol95]). In this theory, one often has to assume that the used mollifier has compact support. We, however, will convolve with the density of a centered normal distribution with a small variance. For primarily fixed \( \varrho > 0 \) consider the density \( k \) of the normal distribution \( N(0, \varrho^2) \) given by

\[
k(x) := \frac{\varrho}{\sqrt{2\pi}} \exp\left(-\frac{\varrho^2 x^2}{2}\right) = \varrho \varphi(\varrho x),
\]

and for a given Lipschitz-continuous test function \( h : \mathbb{R} \to \mathbb{R} \) consider the function \( h_k := h \ast k \) given by

\[
h_k(x) = (h * k)(x) = \int_{\mathbb{R}} h(y)k(x-y)dy = \int_{\mathbb{R}} k(y)h(x-y)dy = (k * h)(x).
\]

By standard results on differentiating integrals depending on a parameter, one can prove the following Proposition.

**Proposition 2.7.5.** For any Lipschitz-continuous function \( h : \mathbb{R} \to \mathbb{R} \), the function \( h_k \) is in \( C^\infty(\mathbb{R}) \) and for each integer \( m \geq 1 \) we have:

(a) \( h_k^{(m)} = h \ast k^{(m)} \)

(b) \( h_k^{(m)} = h' \ast k^{(m-1)} \)

**Corollary 2.7.6.** For each integer \( m \geq 1 \) we have

\[\left\| h_k^{(m)} \right\|_{\infty} \leq \| h' \|_{\infty} \int_{\mathbb{R}} |k^{(m-1)}(y)|dy \leq C_m \varrho^{m-1} \| h' \|_{\infty},\]

where the finite constant \( C_m > 0 \) neither depends on \( \varrho \) nor on \( h \). Furthermore, we may take \( C_1 = 1 \).

**Proof.** By Proposition 2.7.5 (b) for \( x \in \mathbb{R} \) we have

\[|h_k^{(m)}(x)| = \left| \int_{\mathbb{R}} h'(y)k^{(m-1)}(x-y)dy \right| \leq \| h' \|_{\infty} \int_{\mathbb{R}} |k^{(m-1)}(u)|du.
\]

Now observe, that for each integer \( j \geq 1 \) there is a polynomial \( Q_j \) of degree \( j \) and with leading coefficient \( (-1)^j \) such that

\[\varphi^{(j)}(x) = Q_j(x) \varphi(x).
\]
2. Stein’s method for absolutely continuous univariate distributions

and hence, since $k(x) = \varrho \varphi(\varrho x)$,

$$k^{(j)}(u) = \varrho^{j+1} \varphi^{(j)}(\varrho u) = \varrho^j \varrho Q_j(\varrho u) \varphi(\varrho u).$$

Thus, we can conclude, that

$$\int_{\mathbb{R}} \left| k^{(m-1)}(u) \right| du = \varrho^{m-1} \int_{\mathbb{R}} \varrho |Q_{m-1}(\varrho u)| \varphi(\varrho u) du = \varrho^{m-1} \int_{\mathbb{R}} |Q_{m-1}(s)| \varphi(s) ds,$$

so that we can take $C_m := \int_{\mathbb{R}} |Q_{m-1}(s)| \varphi(s) ds.$

The next lemma will be useful.

**Lemma 2.7.7.** Let $\mu$ be a probability measure on $(\mathbb{R}, B)$ with $\int_{\mathbb{R}} |x| d\mu(x) < \infty$. Then, for any Lipschitz-continuous function $h : \mathbb{R} \to \mathbb{R}$ it holds that

$$\left| \int_{\mathbb{R}} h d\mu - \int_{\mathbb{R}} h_k d\mu \right| \leq \frac{\|h'\|_{\infty}}{\varrho}.$$

**Proof.** Since $h$ is Lipschitz with Lipschitz constant $\|h'\|_{\infty}$ we have for each $x \in \mathbb{R}$

$$\left| h_k(x) - h(x) \right| = \left| \int_{\mathbb{R}} (h(x - y) - h(x)) k(y) dy \right|$$

$$\leq \int_{\mathbb{R}} |h(x - y) - h(x)| |k(y) dy| \leq \|h'\|_{\infty} \int_{\mathbb{R}} |y| k(y) dy$$

$$\leq \|h'\|_{\infty} \left( \int_{\mathbb{R}} y^2 k(y) dy \right)^{1/2} = \frac{\|h'\|_{\infty}}{\varrho}$$

and, hence

$$\|h_k - h\|_{\infty} \leq \frac{\|h'\|_{\infty}}{\varrho}.$$ 

Thus,

$$\left| \int_{\mathbb{R}} h d\mu - \int_{\mathbb{R}} h_k d\mu \right| \leq \int_{\mathbb{R}} |h_k(x) - h(x)| d\mu(x)$$

$$\leq \|h_k - h\|_{\infty} \mu(\mathbb{R}) \leq \frac{\|h'\|_{\infty}}{\varrho}.$$ 

□
Now, for $h$ Lipschitz-continuous, by the triangle inequality and Lemma 2.7.7 we can write

$$|\bar{L}_n(h) - \mu_{\frac{1}{2}, \frac{1}{2}}(h)| \leq |\bar{L}_n(h) - \bar{L}_n(h_k)| + |\bar{L}_n(h_k) - \mu_{\frac{1}{2}, \frac{1}{2}}(h_k)|$$

$$+ |\mu_{\frac{1}{2}, \frac{1}{2}}(h_k) - \mu_{\frac{1}{2}, \frac{1}{2}}(h)|$$

$$\leq \frac{2\|h''\|_{\infty}}{\varrho} + |\bar{L}_n(h_k) - \mu_{\frac{1}{2}, \frac{1}{2}}(h_k)|. \quad (2.63)$$

By Corollary 2.7.6 the function $h_k$ has bounded derivatives and thus we can conclude from Theorem 2.7.3 that

$$|\bar{L}_n(h_k) - \mu_{\frac{1}{2}, \frac{1}{2}}(h_k)| \leq K\left(\frac{\|h''\|_{\infty}}{\varrho} + \frac{1}{16n^2} + \frac{\varrho C_2 + \varrho^2 C_3}{16n^2}\right) \quad (2.64)$$

and by (2.63) we obtain

$$|\bar{L}_n(h) - \mu_{\frac{1}{2}, \frac{1}{2}}(h)| \leq \|h''\|_{\infty} \left(K\frac{1}{16n^2} + \frac{\varrho C_2 + \varrho^2 C_3}{16n^2} + \frac{2}{\varrho}\right). \quad (2.65)$$

Choosing $\varrho$ of the order $n^{2/3}$ optimizes the bound in (2.65) and leads to the following theorem.

**Theorem 2.7.8.** For the Wasserstein distance between the expected spectral distribution $\bar{L}_n$ and the semicircle distribution $\mu_{\frac{1}{2}, \frac{1}{2}}$ on $[-1, 1]$ we have

$$d_W(\bar{L}_n, \mu_{\frac{1}{2}, \frac{1}{2}}) = O\left(n^{-2/3}\right).$$

**Remark 2.7.9.** As was already mentioned in Remark 2.7.4, Götze and Tikhomirov [GT05] even obtain a rate of order $n^{-1}$ in the Kolmogorov distance, improving on their former rate of order $n^{-2/3}$ proven in [GT02]. Since the density of the semicircle distribution is bounded, Theorem 2.7.8 immediately yields a rate of order $n^{-1/3}$ in the Kolmogorov distance. This is a well known fact and the proof given in [CS05] for the normal density can easily be adapted to other bounded densities.
2.8. The arcsine law

In this section we present results from the recent preprint [Döb12a] and prove a rate of convergence result for the so-called (first) arcsine law for symmetric random walk on $\mathbb{Z}$. This is possibly best explained in the context of the gambling interpretation of this stochastic process:

Consider a game between two players $A$ and $B$ that consists of consecutive tossings of a fair coin. Each time the coin shows heads, player $A$ has to pay one dollar to player $B$ and conversely, each time the coin shows tails, player $A$ obtains one dollar from player $B$. If we consider the process that gives for each discrete time $n$ the current fortune of player $A$, then by the symmetry of the model one is led to the conjecture, that for $n$ sufficiently large, the relative amount of time, that player $A$ is in the lead, meaning that his or her fortune is positive, should be roughly one-half. The so-called (first) arcsine law states, that this intuition is entirely wrong. In fact, it is more likely that one of the players will lead for nearly all of the time.

Let $(S_k)_{k \geq 0}$ be the symmetric random walk on $\mathbb{Z}$, i.e. we have

$S_k := \sum_{j=1}^k \varepsilon_j$ ($k \geq 0$) for i.i.d. random variables $\varepsilon_1, \varepsilon_2, \ldots$ with $P(\varepsilon_1 = 1) = P(\varepsilon_1 = -1) = \frac{1}{2}$. Letting $X_j := 1_{\{S_{j-1} \geq 0, S_j \geq 0\}}$, $T_m := \sum_{j=1}^{2m} X_j$, $R_m := \frac{1}{m}T_m$ and $W_m := \frac{1}{2m}T_m = \frac{1}{m}R_m$ it is a classical result, first proven by Paul Lévy for Brownian motion, that as $m \to \infty$ we have

$$\mathcal{L}(W_m) \stackrel{D}{\to} \nu_{\frac{1}{2}, \frac{1}{2}},$$

where we recall from Section 2.5 that $\nu := \nu_{\frac{1}{2}, \frac{1}{2}}$ is the special Beta distribution on $[0, 1]$ with parameters $a = b = \frac{1}{2}$, which has density $q(x) := q_{\frac{1}{2}, \frac{1}{2}}(x) = \frac{1}{\pi \sqrt{x(1-x)}} 1_{(0,1)}(x)$. The corresponding distribution function $F$ is given by

$$F(x) = \frac{2}{\pi} \arcsin(\sqrt{x}) = \frac{1}{2} + \frac{1}{\pi} \arcsin(2x - 1),$$

which is why $\nu$ is coined the *arcsine distribution* on $[0, 1]$. A proof of this well-known theorem can be found for example in [Fel68]. Let us recall Stein’s method for $\nu$ from Section 2.5. In this special case the Stein equation corresponding to a given $\nu$-integrable test function $h : \mathbb{R} \to \mathbb{R}$ is given by

$$x(1-x)f'(x) + \left(\frac{1}{2} - x\right)f(x) = h(x) - \nu(h). \quad (2.66)$$

From the theory in Section 2.5 we know that there exists a unique solution $f_h$ to (2.66), defined on $\mathbb{R}$, which is bounded on $[0, 1]$. For $x \in (0,1)$ it is given by
2.8. The arcsine law

\[ f_h(x) = \frac{1}{x(1-x)q(x)} \int_0^x (h(t)-\nu(h))q(t)dt = \frac{-1}{x(1-x)q(x)} \int_x^1 (h(t)-\nu(h))q(t)dt. \]  

(2.67)

Furthermore, \( f_h \) is continuous at 0 and 1 as long as \( h \) is. The following result is a special case of Proposition 2.5.9 from Section 2.5.

**Lemma 2.8.1.** Let \( h : \mathbb{R} \to \mathbb{R} \) be Borel-measurable and \( \nu \)-integrable.

(a) \( h \) is bounded, then \( \|f_h\|_\infty \leq 2\|h - \nu(h)\|_\infty \).

(b) \( h \) is Lipschitz, then \( \|f_h\|_\infty \leq 2\|h'\|_\infty \) and \( \|f'_h\|_\infty \leq C_1\|h\|_\infty \), where the finite constant \( C_1 \) does not depend on \( h \).

(c) \( h \) is twice differentiable with bounded first and second derivative, then \( \|f''_h\|_\infty \leq C_2(\|h'\|_\infty + \|h''\|_\infty ) \), where the finite constant \( C_2 \) does not depend on \( h \).

**Proof.** Noting that \( \frac{1}{2} \) is the median for \( \nu \) with \( q(1/2) = \frac{2}{\pi} \), this follows immediately from Proposition 2.5.9 with \( a = b = \frac{1}{2} \). \( \square \)

**Lemma 2.8.2.** Let \( m \) be a positive integer. Then

(a) \( X_{2j-1} = X_{2j} \) for \( j = 1, \ldots, m \)

(b) \( T_m \) has values in \( 2 \cdot \{0, \ldots, m\} \) and hence \( R_m \) has values in \( \{0, \ldots, m\} \).

(c) Letting \( \tilde{X}_j := 1 - X_j \) we have \( \tilde{X}_j = 1_{\{S_{j-1}\leq 0, S_j \leq 0\}} \) and

\( (\tilde{X}_1, \ldots, \tilde{X}_n) \overset{D}{=} (X_1, \ldots, X_n) \).

**Proof.** We prove \( a \) by induction on \( j \). It is easy to see, that \( X_1 = X_2 \) always holds. Now let \( 1 \leq j \leq m - 1 \). Then we have \( X_{2j-1} = X_{2j} \) by the induction hypothesis.

Suppose, that \( X_{2j-1} = X_{2j} = 1 \). If \( S_{2j} = 0 \), then the claim \( X_{2j+1} = X_{2j+2} \) follows in the same manner as \( X_1 = X_2 \). If \( S_{2j} > 0 \), then necessarily \( S_{2j} \geq 2 \), yielding \( S_{2j+1} > 0 \) and \( S_{2j+2} \geq 0 \). Hence, \( X_{2j+1} = X_{2(j+1)} = 1 \) in this case. If, contrarily, \( X_{2j-1} = X_{2j} = 0 \), the proof is similar.

Assertion \( b \) follows immediately from \( a \). The first assertion from \( c \) is clear since either both, \( S_{2j-1} \) and \( S_{2j} \), are nonnegative or nonpositive. Now, observe, that there is a (measurable) function \( f \) such that \( (X_1, \ldots, X_n) = f(S_1, \ldots, S_n) \).

Since \( \tilde{X}_j = 1_{\{-S_{j-1} \geq 0, -S_j \leq 0\}} \) and \( (S_1, \ldots, S_n) \overset{D}{=} (-S_1, \ldots, -S_n) \) by symmetry, we have

\( (\tilde{X}_1, \ldots, \tilde{X}_n) = f(-S_1, \ldots, -S_n) \overset{D}{=} f(S_1, \ldots, S_n) = (X_1, \ldots, X_n) \). \( \square \)
The proof of the following well-known theorem can be found for example in [Fel68].

**Theorem 2.8.3** (Chung-Feller Theorem). Let \( m \) be a positive integer. Then, for each \( 0 \leq k \leq m \) we have

\[
P(R_m = k) = P(T_m = 2k) = u_{2k}u_{2m-2k},
\]

where \( u_0 := 1 \) and \( u_{2j} := (\binom{2j}{j})2^{-j} \) for \( j \geq 1 \) denotes the probability that the symmetric random walk returns to zero at time \( 2j \).

Thus, by Theorem 2.8.3 the probability mass function \( p : \mathbb{Z} \to \mathbb{R} \) corresponding to \( R_m \) is given by \( p(k) = 0 \) for \( k \in \mathbb{Z} \setminus \{0, \ldots, m\} \) and by

\[
p(k) = u_{2k}u_{2m-2k} = 2^{-2m}\binom{2k}{k}\binom{2m-2k}{m-k},
\]

for \( 0 \leq k \leq m \).

In the following, we review the recent adoption of the so-called *density approach* for absolutely continuous distributions (see Subsection 2.1.2 as well as, e.g. [SDHR04], [EL10], [CS11] and [CGS11]) to discrete distributions on the integers, which was done in [GR12] and also in [LS11a]. For reasons of simplicity we restrict ourselves to the case of finite integer intervals.

A finite integer interval is a set \( I \) of the form \( I = [a,b] \cap \mathbb{Z} \) for some integers \( a \leq b \). Given a probability mass function \( p : \mathbb{Z} \to \mathbb{R} \) with \( p(k) > 0 \) for \( k \in I \) and \( p(k) = 0 \) for \( k \in \mathbb{Z} \setminus I \), we consider the function \( \psi : I \to \mathbb{R} \) given by the formula

\[
\psi(k) := \frac{\Delta p(k)}{p(k)},
\]

where for a function \( f \) on the integers \( \Delta f(k) := f(k+1) - f(k) \) denotes the forward difference operator. Note that by definition always \( \psi(b) = -1 \) if \( I = [a,b] \cap \mathbb{Z} \). For such a probability mass function \( p \) with support a finite integer interval \( I = [a,b] \cap \mathbb{Z} \), let \( \mathcal{F}(p) \) denote the class of all real-valued functions \( f \) on \( \mathbb{Z} \) such that \( f(a-1) = 0 \). The following result is a special case of Proposition 2.1 of [GR12].

**Proposition 2.8.4.** Let \( Z \) be a \( \mathbb{Z} \)-valued random variable with probability mass function \( p \) which is supported on the finite integer interval \( I = [a,b] \cap \mathbb{Z} \) and is positive there. Then, a given random variable \( X \) with support \( I \) has the probability mass function \( p \) if and only if for all \( f \in \mathcal{F}(p) \) it holds that

\[
E[\Delta f(X-1) + \psi(X)f(X)] = 0.
\]
The next result, a version of Corollary 2.1 from [GR12], yields various other Stein characterizations for the distribution corresponding to \( p \) from Proposition 2.8.4.

**Corollary 2.8.5.** Let \( Z \) be a \( \mathbb{Z} \)-valued random variable with probability mass function \( p \) which is supported on the finite integer interval \( I = [a, b] \cap \mathbb{Z} \) and is positive there. Let \( c : [a - 1, b] \cap \mathbb{Z} \to \mathbb{R} \) be a function with \( c(k) \neq 0 \) for all \( k \in I \). Then, in order that a given random variable \( X \) with support \( I \) has the probability mass function \( p \) it is necessary and sufficient that for all functions \( f \in \mathcal{F}(p) \) we have

\[
E \left[ c(X - 1) \Delta f(X - 1) + \left[ c(X) \psi(X) + \Delta c(X - 1) \right] f(x) \right] = 0. \tag{2.71}
\]

**Remark 2.8.6.** Letting \( \gamma(k) := c(k)\psi(k) + \Delta c(k - 1) \) we see that \( c \) satisfies the difference equation \( \Delta c(k - 1) = \gamma(k) - c(k)\psi(k) \). This exactly corresponds to the differential equation (2.17) \( \eta'(x) = \gamma(x) - \eta(x)\psi(x) \) from Section 2.4, where \( \psi(x) := \frac{p'(x)}{p(x)} \) is the logarithmic derivative of the density \( p \). So also in this respect, there is a strong analogy between the absolutely continuous and the discrete case.

Now, with the abstract results at hand, we return to the concrete distribution of \( R_m \) which has probability mass function \( p \) supported on \( I := [0, m] \cap \mathbb{Z} \) and which is given by (2.68). Using the relation

\[
\binom{2j}{j} = \frac{4j - 2}{j} \binom{2j - 2}{j - 1} \quad \text{for } j \geq 1
\]

it can easily be checked, that in this case \( \psi \) is given by

\[
\psi(k) = \frac{2k - m + 1}{(k + 1)(2(m - k) - 1)}, \quad 0 \leq k \leq m. \tag{2.72}
\]

This motivates the definition \( c(k) := (k + 1)(2(m - k) - 1) \) for \( k = -1, 0, \ldots, m \). Note, that \( c(k) \neq 0 \) for \( k \in I \). These observations lead to the following lemma.

**Lemma 2.8.7.** Let \( p \) be the probability mass function of \( R_m \) as given by (2.68). A random variable \( X \) with support \( I = [0, m] \cap \mathbb{Z} \) has probability mass function \( p \) if and only if for all \( f \in \mathcal{F}(p) \) it holds that

\[
E \left[ X \left( (m - X) + \frac{1}{2} \right) \Delta f(X - 1) + \left( \frac{m}{2} - X \right) f(x) \right] = 0. \tag{2.73}
\]

**Proof.** This follows from Corollary 2.8.5, since by the definition of \( c \) we have

\[
c(k)\psi(k) + \Delta c(k - 1) = 2\left( \frac{m}{2} - k \right)
\]
2. Stein’s method for absolutely continuous univariate distributions

and

\[ c(k - 1) = k(2(m - k) + 1) = 2k \left( m - k + \frac{1}{2} \right). \]

In the following we will use the above results to prove a rate of convergence in the Wasserstein distance for the arcsine law. Recall, that for two distributions \( \mu_1 \) and \( \mu_2 \) on \((\mathbb{R}, B)\), whose first moments exist, the Wasserstein distance is given by

\[
d_W(\mu_1, \mu_2) := \sup_{h \in \text{Lip}(1)} \left| \int \mathbb{R} h \, d\mu_1 - \int \mathbb{R} h \, d\mu_2 \right|
\]

and for two real-valued random variables \( X \) and \( Y \) one defines

\[
d_W(X, Y) := d_W(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{h \in \text{Lip}(1)} \left| E[h(X)] - E[h(Y)] \right|
\]

where \( \text{Lip}(1) \) denotes the class of all Lipschitz-continuous functions \( h \) on \( \mathbb{R} \) with minimal Lipschitz constant \( \|h'\|_\infty \leq 1 \). It is known, that on the space of probability measures with existing first moment, convergence in the Wasserstein distance is stronger than weak convergence.

**Theorem 2.8.8.** There exists a finite constant \( C > 0 \) such that for each positive integer \( m \) we have

\[
d_W(\mathcal{L}(W_m), \nu) \leq \frac{C}{m}.
\]

**Remark 2.8.9.** To the best of my knowledge this is the first result that gives a rate of convergence of order \( m^{-1} \) for the arcsine law. The restriction to even times \( n = 2m \) is immaterial and only for convenience, since the formula from the Chung-Feller Theorem only holds for these times. Since for odd times \( n = 2m + 1 \)

\[
\left| \frac{1}{2m+1} \sum_{j=1}^{2m+1} X_j - W_m \right| \leq \frac{2}{2m+1},
\]

the same rate of convergence also holds for the whole sequence of positive times of the random walk.

It may be seen from the proof of Theorem 2.8.8, that the constant \( C \) can be made explicit in terms of the constant \( C_1 \) from Lemma 2.8.1.
2.8. The arcsine law

**Proof of Theorem 2.8.8.** The proof is very similar to the proof of Theorem 3.1 from [GR12]. Using the notation from [GR12], for a function $f$ and $y > 0$ let

$$\Delta_y f(x) := f(x + y) - f(x).$$

We also write $W := W_m$ and $R := R_m$. Let $h \in \text{Lip}(1)$ be fixed and let $f := f_h$ be the corresponding solution to the Stein equation (2.66) given by (2.67) for $x \in (0, 1)$ but which we set equal to zero for $x \in \mathbb{R} \setminus [0, 1]$. Consider the function $g(x) := f(x/m)$ which is zero on $\mathbb{R} \setminus [0, m]$. Then by Lemma 2.8.7 and upon dividing by $m$ in (2.73) we obtain

$$0 = \frac{1}{m} E \left[ R \left( (m - R) + \frac{1}{2} \right) \Delta g(R - 1) + \left( \frac{m}{2} - R \right) g(R) \right]$$

$$= E \left[ mW \left( (1 - W) + \frac{1}{2m} \right) \Delta_{1/m} f \left( W - \frac{1}{m} \right) + \left( \frac{1}{2} - W \right) f(W) \right].$$

Inserting this into the Stein identity resulting from the Stein equation (2.66) we obtain

$$E[h(W)] - \nu(h)$$

$$= E \left[ W(1 - W)f'(W) + \left( \frac{1}{2} - W \right) f(W) \right]$$

$$= E \left[ W(1 - W)f'(W) - mW \left( (1 - W) + \frac{1}{2m} \right) \Delta_{1/m} f \left( W - \frac{1}{m} \right) \right]$$

$$= E \left[ W(1 - W)f'(W) - mW(1 - W) \Delta_{1/m} f \left( W - \frac{1}{m} \right) \right] + E_1, \quad (2.74)$$

with

$$|E_1| = \frac{1}{2} \left| E \left[ W \Delta_{1/m} f \left( W - \frac{1}{m} \right) \right] \right| \leq \frac{1}{2m} \|f'\|_{\infty} E[W] = \frac{1}{4m} \|f'\|_{\infty} \leq \frac{C_1}{4m} \quad (2.75)$$

by Lemma 2.8.1 (b) and by $E[W] = 1/2$, which follows for example from Lemma 2.8.2 (c) by symmetry. Using the fundamental theorem of calculus, we rewrite the remaining term in (2.74) as
2. Stein’s method for absolutely continuous univariate distributions

\[ E \left[ W(1 - W)f'(W) - mW(1 - W)\Delta_{1/m}(W - \frac{1}{m}) \right] \]

\[ = E \left[ W(1 - W) \left( f'(W) - m \int_{W - \frac{1}{m}}^{W} f'(t) dt \right) \right] \]

\[ = mE \left[ \int_{W - \frac{1}{m}}^{W} W(1 - W) \left( f'(W) - f'(t) \right) dt \right] \]

\[ = mE \left[ \int_{W - \frac{1}{m}}^{W} \left( W(1 - W)f'(W) - t(1 - t)f'(t) \right) dt \right] + E_2 \quad (2.76) \]

where

\[ E_2 := -mE \left[ \int_{W - \frac{1}{m}}^{W} \left( W(1 - W) - t(1 - t) \right) f'(t) dt \right] \]

and hence, again by the fundamental theorem of calculus,

\[ |E_2| = m \left| E \left[ \int_{W - \frac{1}{m}}^{W} f'(t) \int_t^{W} (1 - 2s) ds dt \right] \right| \]

\[ \leq m \left( 1 + \frac{2}{m} \right) \| f' \|_{\infty} E \left[ \int_{W - \frac{1}{m}}^{W} (W - t) dt \right] \]

\[ = (m + 2) \| f' \|_{\infty} \int_0^{\frac{1}{m}} u du \]

\[ = \frac{m + 2}{2m^2} \| f' \|_{\infty} \leq C_1 \frac{m + 2}{2m^2} \quad (2.77) \]

where we have used Lemma 2.8.1 for the last step and the inequality \(|1 - 2s| \leq (1 + \frac{2}{m})\) for relevant values of \(s\) for the first step (recall that \(W \in [0, 1]\)).

It remains to deal with the first expectation in (2.76). Since \(f = f_h\) solves the Stein equation (2.66) and by the fundamental theorem of calculus for Lebesgue integration, we obtain
2.8. The arcsine law

\[ mE \left[ \int_{W - \frac{1}{m}}^{W} \left( W(1 - W) f'(W) - t(1 - t) f'(t) \right) dt \right] \]

\[ = mE \left[ \int_{W - \frac{1}{m}}^{W} \left( h(W) - \nu(h) - \left( \frac{1}{2} - W \right) f(W) - h(t) + \nu(h) + \left( \frac{1}{2} - t \right) f(t) \right) dt \right] \]

\[ = mE \left[ \int_{W - \frac{1}{m}}^{W} \left( h(W) - h(t) + \left( W - \frac{1}{2} \right) f(W) - \left( t - \frac{1}{2} \right) f(t) \right) dt \right] \]

\[ = mE \left[ \int_{W - \frac{1}{m}}^{W} \left( \int_{t}^{W} h'(s) ds + \int_{t}^{W} \left( f(s) + \left( s - \frac{1}{2} \right) f'(s) \right) ds \right) dt \right] \]

The inner integrals from (2.78) are bounded separately. As to the first one,

\[ \left| mE \left[ \int_{W - \frac{1}{m}}^{W} \int_{t}^{W} h'(s) ds dt \right] \right| \leq m \| h' \|_{\infty} E \left[ \int_{W - \frac{1}{m}}^{W} \left( W - t \right) dt \right] = \frac{1}{2m} \| h' \|_{\infty} \leq \frac{1}{2m}. \]

(2.79)

For the second one, since \(| \frac{1}{2} - s | \leq \frac{1}{m} + \frac{1}{2} \leq 3/2\) for the relevant values of \( s \), we have

\[ m \left| E \left[ \int_{W - \frac{1}{m}}^{W} \int_{t}^{W} \left( f(s) + \left( s - \frac{1}{2} \right) f'(s) \right) ds dt \right] \right| \]

\[ \leq m \left( \| f \|_{\infty} + \frac{3}{2} \| f' \|_{\infty} \right) \left[ \int_{W - \frac{1}{m}}^{W} \int_{t}^{W} ds dt \right] \]

\[ = m \left( \| f \|_{\infty} + \frac{3}{2} \| f' \|_{\infty} \right) \left[ \int_{W - \frac{1}{m}}^{W} \left( W - t \right) dt \right] \]

\[ = m \left( \| f \|_{\infty} + \frac{3}{2} \| f' \|_{\infty} \right) \frac{1}{2m^{2}} \]

\[ \leq \left( 2 + \frac{3}{2} C_{1} \right) \| h' \|_{\infty} \frac{1}{2m} \leq \frac{4 + 3C_{1}}{4m}, \]

(2.80)

where we have used Lemma 2.8.1 for the next to last inequality. Since \( h \in \text{Lip}(1) \) was arbitrary, the conclusion of the theorem follows from (2.75), (2.77), (2.79) and (2.80).\[ \square \]
3. Stein’s method for multivariate normal approximation with applications to the classical compact groups

In this final chapter we turn to multivariate normal approximation by Stein’s method. Following the presentation in [Mec09], in Section 3.1 we provide a Stein characterization and discuss properties of the solution to Stein’s equation for arbitrary centered multivariate normal distributions. We also present the exchangeable pairs approach in the framework of multivariate normal approximation, which goes back to the papers [CM08], [RR09] and [Mec09]. For our applications in Section 3.2 we prove a complex version (see Theorem 3.1.17) of the so-called *infinitesimal version of exchangeable pairs* which goes back to the paper [Ste95] and which was further developed in [Mec06], [CM08] and in [Mec09] for real-valued random variables and random vectors in $\mathbb{R}^d$, respectively. With possible future applications in mind, we also state and prove a complex version of the standard plug-in theorem for a fixed exchangeable pair $(W, W')$ of random vectors.

In Section 3.2 the theory from Section 3.1 is applied to study the vector of traces of various powers of a Haar distributed element from one of the classical, connected compact Lie groups. More precisely, we prove a quantitative version of the Diaconis-Shahshahani theorem from [DS94], which states that the vector of traces of powers is asymptotically normal and that its entries are asymptotically independent. Finally, the result on the vector of traces of powers is applied to prove a rate of convergence for the Gaussian fluctuations of suitable linear statistics of the eigenvalues of a Haar distributed unitary matrix as the matrix dimension tends to infinity.
3. Multivariate normal approximation and the classical compact groups

3.1. Stein’s method for multivariate normal distributions

At the end of Section 1.2, for the multivariate standard normal distribution, we already found a Stein equation (1.16) and a corresponding solution \( f \) given by (1.18) in quite explicit form. The goal of the present section is to give an account of Stein’s method for the general centered \( d \)-dimensional normal distribution \( N_d(0, \Sigma) \), where \( \Sigma \in \mathbb{R}^{d \times d} \) is a given positive semidefinite matrix. Of course, the assumption, that the expected value equals zero is only for convenience.

3.1.1. The Stein equation and its solution

In order to find a suitable Stein equation and a formula for the solution, we use the existing results for the \( d \)-dimensional standard normal distribution \( N_d(0, I_d) \). In particular, from Proposition 1.2.1 and since the Stein operator for \( N_d(0, I_d) \) is given by

\[
\Delta - \langle x, \nabla \rangle,
\]

we know that

\[
E[\Delta g(Z) - \langle Z, \nabla g(Z) \rangle] = 0 \tag{3.1}
\]

for all \( g \in C^2_c(\mathbb{R}^d) \). Here and in what follows \( Z \sim N_d(0, I_d) \). Let us first assume that \( \Sigma \) is in fact positive definite and let \( f \in C^2_c(\mathbb{R}^d) \). Then the function \( g \) with \( g(x) := f(\Sigma^{1/2}x) \) is also in \( C^2_c(\mathbb{R}^d) \) and thus (3.1) holds. By the chain rule formula we have

\[
\nabla g(x) = \Sigma^{1/2} \nabla f(\Sigma^{1/2}x),
\]

\[
\text{Hess } g(x) = D(\nabla g)(x) = \Sigma^{1/2} D(\nabla f)(\Sigma^{1/2}x) \Sigma^{1/2} = \Sigma^{1/2} \text{Hess } f(\Sigma^{1/2}x) \Sigma^{1/2}
\]

and, hence,

\[
\Delta g(x) = \text{Tr}(\text{Hess } g(x)) = \text{Tr} \left( \Sigma^{1/2} \text{Hess } f(\Sigma^{1/2}x) \Sigma^{1/2} \right)
\]

\[
= \text{Tr} \left( \Sigma \text{Hess } f(\Sigma^{1/2}x) \right) = \langle \text{Hess } f(\Sigma^{1/2}x), \Sigma \rangle_{\text{H.S.}}, \tag{3.3}
\]

where the \( \text{Hilbert-Schmidt inner product} \) of two matrices \( A, B \in \mathbb{R}^{d \times d} \) is defined by
3.1. Stein’s method for multivariate normal distributions

\[ \langle A, B \rangle_{\text{H.S.}} := \text{Tr}(AB^T) = \text{Tr}(BA^T) = \text{Tr}(B^TA) = \sum_{i,j=1}^{d} a_{ij}b_{ij}. \]

Thus, \( \langle \cdot, \cdot \rangle_{\text{H.S.}} \) is just the standard inner product on \( \mathbb{R}^{d \times d} \cong \mathbb{R}^d \). The corresponding norm will be denoted by \( \| \cdot \|_{\text{H.S.}} \).

So, letting \( Z_\Sigma := \Sigma^{1/2}Z \sim N_d(0, \Sigma) \), we conclude from (3.1), (3.2) and (3.3) that

\[
0 = E \left[ \Delta g(Z) - \langle Z, \nabla g(Z) \rangle \right] = E \left[ \langle \text{Hess} f(\Sigma^{1/2}Z), \Sigma \rangle_{\text{H.S.}} - Z^T \Sigma^{1/2} \nabla f(\Sigma^{1/2}Z) \right]
\]

where we have used that \( Z^T \Sigma^{1/2} = (\Sigma^{1/2}Z)^T = Z^T \Sigma \) by the symmetry of \( \Sigma^{1/2} \).

Hence, a Stein identity for \( Z_\Sigma \) is given by

\[
E \left[ \langle \text{Hess} f(Z_\Sigma), \Sigma \rangle_{\text{H.S.}} - Z^T_\Sigma \nabla f(Z_\Sigma) \right] = 0 \quad (3.4)
\]

for \( f \in C^2_c(\mathbb{R}^d) \). By the Stein identity (3.4) for a smooth enough test function \( h : \mathbb{R}^d \rightarrow \mathbb{R} \), one is led to consider the following Stein equation:

\[
\langle \text{Hess} f(x), \Sigma \rangle_{\text{H.S.}} - x^T \nabla f(x) = h(x) - E[h(Z_\Sigma)] \quad (3.5)
\]

For a given function \( f \in C^2_c(\mathbb{R}^d) \), the function \( g(x) := f(\Sigma^{1/2}x) \) is also in \( C^2_c(\mathbb{R}^d) \) and if \( g \) satisfies the equation

\[
\Delta g(x) - x^T \nabla g(x) = h(\Sigma^{1/2}x) - E[h(Z_\Sigma)], \quad (3.6)
\]

then by (3.3) and (3.2) the function \( f \) satisfies (3.5). From (1.18) we know that \( g \) should be given by

\[
g(x) = - \int_0^1 \frac{1}{2t} \left( E \left[ h(\sqrt{t}\Sigma^{1/2}x + \sqrt{1-t}Z_\Sigma) \right] - E[h(Z_\Sigma)] \right) dt
\]

and hence letting \( f(x) = g(\Sigma^{-1/2}x) \), for the Stein equation (3.5) we have the candidate solution \( f_h \) with

\[
f_h(x) = - \int_0^1 \frac{1}{2t} \left( E \left[ h(\sqrt{t}x + \sqrt{1-t}Z_\Sigma) \right] - E[h(Z_\Sigma)] \right) dt. \quad (3.7)
\]
3. Multivariate normal approximation and the classical compact groups

Although the derivation of this formula for \( f_h \) used the regularity of \( \Sigma \) we will see, that even for singular \( \Sigma \) formula (3.7) gives the right solution to (3.5).

We now turn to properties of the function \( f_h \) and show in particular, that it is well-defined if \( h \) is Lipschitz-continuous.

\textbf{Lemma 3.1.1.} If the function \( h : \mathbb{R}^d \to \mathbb{R} \) is Lipschitz-continuous with constant \( M > 0 \), then the function \( f_h \) given by (3.7) is well-defined.

\textbf{Proof.} Let \( x \in \mathbb{R}^d \) be fixed. Then,

\[
\int_0^1 \frac{1}{2t} \mathbb{E} \left[ \left| h(\sqrt{t}x + \sqrt{1-t}Z_\Sigma) - h(Z_\Sigma) \right| \right] dt \\
\leq M \int_0^1 \frac{1}{2t} \mathbb{E} \left[ \left\| (\sqrt{t}x + \sqrt{1-t}Z_\Sigma) - Z_\Sigma \right\|_2 \right] \\
\leq M \| x \|_2 \int_0^1 \frac{1}{2\sqrt{t}} dt + LE \left[ \| Z_\Sigma \|_2 \right] \int_0^1 \frac{1 - \sqrt{1-t}}{2t} dt \\
< \infty
\]

since \( \int_0^1 \frac{1}{2\sqrt{t}} dt = 1 \) and \( \int_0^1 \frac{1 - \sqrt{1-t}}{2t} dt < \infty \) by continuity of the integrand at \( t = 0 \). Thus, the existence of \( f_h(x) \) follows from the Fubini-Tonelli theorem. \( \square \)

\textbf{Corollary 3.1.2.} If the function \( h : \mathbb{R}^d \to \mathbb{R} \) is differentiable with \( \sup_{x \in \mathbb{R}^d} \| Dh(x) \|_\text{op} < \infty \), then the function \( f_h \) given by (3.7) is well-defined.

\textbf{Proof.} By Lemma 3.1.1 it suffices to show that \( h \) is Lipschitz-continuous with constant \( M := \sup_{x \in \mathbb{R}^d} \| Dh(x) \|_\text{op} \). To this end fix \( x, y \in \mathbb{R}^d \) and define the function \( g : [0, 1] \to \mathbb{R} \) by

\[ g(t) := h(x + t(y - x)). \]

Then \( g \) is differentiable with

\[ |g'(t)| = |Dh(x + t(y - x))(y - x)| \leq \| Dh(x + t(y - x)) \|_\text{op} \| y - x \|_2 \leq M \| y - x \|_2 \]

for all \( t \in [0, 1] \). By a well-known theorem from measure and integration theory this implies, that \( g' \) is Lebesgue-integrable over \([0, 1]\) and

\[ g(1) - g(0) = \int_0^1 g'(t)dt. \]
3.1. Stein’s method for multivariate normal distributions

Hence, we obtain

\[ |h(y) - h(x)| = |g(1) - g(0)| \leq \int_0^1 |g'(t)| dt \leq M \|y - x\|_2. \]

Since \(x\) and \(y\) were arbitrary, the result follows.

\[ \square \]

In the following we will show, that the function \(f_h\) given by (3.7) actually solves the Stein equation (3.5), if the test function \(h\) is twice differentiable with bounded first and second derivative. The proof of the following lemma, which goes back to Meckes \cite{Mec09}, is direct and does not appeal to the theory of operator semigroups as discussed in Section 1.2.

**Lemma 3.1.3.** Let \(h : \mathbb{R}^d \to \mathbb{R}\) be a twice differentiable test function, such that \(\sup_{x \in \mathbb{R}^d} \|Dh(x)\|_{op} < \infty\) and \(\sup_{x \in \mathbb{R}^d} \|D^2h(x)\| < \infty\) in any of the equivalent norms. Then, the function \(f_h\) given by (3.7) solves the Stein equation (3.5).

**Proof.** We follow the lines of proof of Lemma 1 item 3 in \cite{Mec09}, but present a more detailed account of the proof. In particular the “integration by parts” argument, mentioned there, will be explained in detail.

Fix \(x \in \mathbb{R}^d\). For \(t \in (0, 1)\) let

\[ Z_{\Sigma, x, t} := \sqrt{tx} + \sqrt{1-t}Z_{\Sigma}. \]

By the fundamental theorem of calculus and the properties of \(h\), we have

\[ h(x) - E[h(Z_{\Sigma})] = \int_0^1 \frac{d}{dt} E[h(Z_{\Sigma, x, t})] dt = \int_0^1 E\left[ \frac{d}{dt} h(Z_{\Sigma, x, t}) \right] dt \]

\[ = \int_0^1 E\left[ Dh(Z_{\Sigma, x, t}) \left( \frac{x}{2\sqrt{t}} - \frac{Z_{\Sigma}}{2\sqrt{1-t}} \right) \right] dt \]

\[ = \int_0^1 \frac{1}{2\sqrt{t}} E\left[ \langle x, \nabla h(Z_{\Sigma, x, t}) \rangle \right] dt - \int_0^1 \frac{1}{2\sqrt{1-t}} E\left[ \langle Z_{\Sigma}, \nabla h(Z_{\Sigma, x, t}) \rangle \right] dt. \]

Let \(C := (c_{ij})_{i,j=1,...,d} := \Sigma^{1/2}\). Then we can write

101
3. Multivariate normal approximation and the classical compact groups

\[
E \left[ \langle Z_\Sigma, \nabla h(Z_{\Sigma,x,t}) \rangle \right] = E \left[ \langle CZ, \nabla h(Z_{\Sigma,x,t}) \rangle \right]
\]

\[
=E \sum_{i=1}^{d} E \left[ \sum_{j=1}^{d} c_{ij} Z_j \frac{\partial h}{\partial y_i} (Z_{\Sigma,x,t}) \right]
\]

\[
= \sum_{i=1}^{d} \sum_{j=1}^{d} c_{ij} E \left[ Z_j \frac{\partial h}{\partial y_i} \left( \sqrt{tx} + \sqrt{1-t} \sum_{k=1}^{d} (\sum_{l=1}^{d} c_{kl} e_k) Z_l \right) \right]
\]

\[
= \sum_{i=1}^{d} \sum_{j=1}^{d} c_{ij} \frac{\partial^2 h}{\partial y_k \partial y_i} (Z_{\Sigma,x,t})
\]

(3.9)

where \( e_k \) denotes the \( k \)-th standard unit basis vector in \( \mathbb{R}^d \). Now, applying for each fixed index \( j \in \{1, \ldots, d\} \) the one-dimensional Stein identity given by Proposition 1.1.1 to the univariate standard normal variable \( Z_j \), we obtain by the independence of \( Z_1, \ldots, Z_d \) that

\[
E \left[ Z_j \frac{\partial h}{\partial y_i} \left( \sqrt{tx} + \sqrt{1-t} \sum_{l=1}^{d} (\sum_{k=1}^{d} c_{kl} e_k) Z_l \right) \right] = \sqrt{1-t} \sum_{k=1}^{d} c_{kj} E \left[ \frac{\partial^2 h}{\partial y_k \partial y_i} (Z_{\Sigma,x,t}) \right].
\]

Hence, using \( c_{kj} = c_{jk} \) we conclude from (3.9) that with \( \Sigma = (\sigma_{ij})_{i,j=1,\ldots,d} \)

\[
E \left[ (Z_\Sigma, \nabla h(Z_{\Sigma,x,t})) \right] = \sqrt{1-t} \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} c_{ij} c_{kj} E \left[ \frac{\partial^2 h}{\partial y_k \partial y_i} (Z_{\Sigma,x,t}) \right]
\]

\[
= \sqrt{1-t} \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} \sigma_{ik} E \left[ \frac{\partial^2 h}{\partial y_k \partial y_i} (Z_{\Sigma,x,t}) \right]
\]

\[
= \sqrt{1-t} \text{Tr} \left( \Sigma E [\text{Hess} h(Z_{\Sigma,x,t})] \right)
\]

\[
= \sqrt{1-t} E \left[ \langle \Sigma, \text{Hess} h(Z_{\Sigma,x,t}) \rangle_{\text{H.S.}} \right].
\]

(3.10)

Thus we arrive at the representation

\[
h(x) - E[h(Z)] = \int_0^1 \frac{1}{2 \sqrt{t}} E \left[ \langle x, \nabla h(Z_{\Sigma,x,t}) \rangle \right] dt - \frac{1}{2} \int_0^1 E \left[ \langle \Sigma, \text{Hess} h(Z_{\Sigma,x,t}) \rangle_{\text{H.S.}} \right] dt.
\]

(3.11)
3.1. Stein’s method for multivariate normal distributions

Now, by the conditions on \( h \) it follows that for \( j, k \in \{1, \ldots, d\} \)

\[
\frac{\partial f_h}{\partial x_j}(x) = -\frac{1}{2} \int_0^1 \frac{1}{\sqrt{t}} E \left[ \frac{\partial h}{\partial y_j}(Z_{\Sigma, x, t}) \right] dt \tag{3.12}
\]

and

\[
\frac{\partial^2 f_h}{\partial x_k \partial x_j}(x) = -\frac{1}{2} \int_0^1 E \left[ \frac{\partial^2 h}{\partial y_k \partial y_j}(Z_{\Sigma, x, t}) \right] dt. \tag{3.13}
\]

By the symmetry of \( \Sigma \) this finally implies that

\[
\langle \text{Hess } f_h(x), \Sigma \rangle_{\text{H.S.}} - x^T \nabla f_h(x) = \text{Tr} \left( \Sigma \text{Hess } f_h(x) \right) - x^T \nabla f_h(x)
\]

\[
= \sum_{j, k=1}^d \sigma_{jk} \frac{\partial^2 f_h}{\partial x_k \partial x_j}(x) - \sum_{j=1}^d x_j \frac{\partial f_h}{\partial x_j}(x)
\]

\[
= -\frac{1}{2} \sum_{j, k=1}^d \sigma_{jk} \int_0^1 E \left[ \frac{\partial^2 h}{\partial y_k \partial y_j}(Z_{\Sigma, x, t}) \right] dt + \frac{1}{2} \sum_{j=1}^d x_j \int_0^1 \frac{1}{\sqrt{t}} E \left[ \frac{\partial h}{\partial y_j}(Z_{\Sigma, x, t}) \right] dt
\]

\[
= -\frac{1}{2} \int_0^1 E \left[ \text{Tr} \left( \Sigma \text{Hess } h(Z_{\Sigma, x, t}) \right) \right] dt + \frac{1}{2} \int_0^1 \frac{1}{\sqrt{t}} E \left[ \langle x, \nabla (Z_{\Sigma, x, t}) \rangle \right] dt
\]

\[
= \int_0^1 \frac{1}{2\sqrt{t}} E \left[ \langle x, \nabla (Z_{\Sigma, x, t}) \rangle \right] dt - \frac{1}{2} \int_0^1 E \left[ \langle \Sigma, \text{Hess } h(Z_{\Sigma, x, t}) \rangle_{\text{H.S.}} \right] dt. \tag{3.14}
\]

The claim now follows by identifying (3.11) and (3.14).

Next, we would like to present concrete bounds on the solution \( f_h \) and its derivatives in terms of bounds on the test function \( h \) and its derivatives. To do so, we first have to choose suitable quantities, which measure the size of a function or of its derivatives, respectively. The most common choice is to consider for a \( k \)-times differentiable function \( h : \mathbb{R}^d \rightarrow \mathbb{R} \) the quantities

\[
|h|_r := \sup_{1 \leq i_1, \ldots, i_r \leq d} \| \frac{\partial^r h}{\partial x_{i_1} \cdots \partial x_{i_r}} \|_\infty, \quad r = 0, \ldots, k.
\]

Bounds on the normal approximation on a given random vector in terms of these quantities are for example given in [RR09] and in [GR96]. It was noticed in [CM08] and in [Mec09] that the following coordinate-free quantities often yield better rates of convergence with respect to the dimension \( d \). To introduce them,
3. Multivariate normal approximation and the classical compact groups

we have to recall some notions and concepts.

For $A \in \mathbb{R}^{d \times d}$ let $\|A\|_{\text{op}}$ denote the operator norm induced by the Euclidean norm, i.e., $\|A\|_{\text{op}} := \sup\{\|Ax\|_2 : \|x\|_2 = 1\}$. More generally, for a $k$-multilinear form $\psi : (\mathbb{R}^d)^k \to \mathbb{R}$ define the operator norm

$$
\|\psi\|_{\text{op}} := \sup \left\{ |\psi(u_1, \ldots, u_k)| : u_j \in \mathbb{R}^d, \|u_j\|_2 = 1, j = 1, \ldots, k \right\}.
$$

For a function $h : \mathbb{R}^d \to \mathbb{R}$ define its minimum Lipschitz constant $M_1(h)$ by

$$
M_1(h) := \sup_{x \neq y} \frac{\|h(x) - h(y)\|}{\|x - y\|_2} \in [0, \infty) \cup \{\infty\}.
$$

If $h$ is differentiable, then $M_1(h) = \sup_{x \in \mathbb{R}^d} \|Dh(x)\|_{\text{op}}$. More generally, for $k \geq 1$ and a $(k-1)$-times differentiable function $h : \mathbb{R}^d \to \mathbb{R}$ let

$$
M_k(h) := \sup_{x \neq y} \frac{\|D^{k-1}h(x) - D^{k-1}h(y)\|_{\text{op}}}{\|x - y\|_2},
$$

viewing the $(k-1)$-th derivative of $h$ at any point as a $(k-1)$-multilinear form. Then, if $h$ is actually $k$-times differentiable, we have $M_k(h) = \sup_{x \in \mathbb{R}^d} \|D^k h(x)\|_{\text{op}}$. Having in mind this identity, we define $M_0(h) := \|h\|_{\infty}$. Additionally, following [CM08] and [Mec09], for $k = 2$ we define

$$
\tilde{M}_2(h) := \sup_{x \in \mathbb{R}^d} \|\text{Hess } h(x)\|_{\text{H.S.}}.
$$

The next lemma gives bounds on the solution $f_h$ and its derivatives in terms of the quantities $M_k(h)$ and $\tilde{M}_2(h)$. It is a more general version of Lemma 2 in [Mec09], where the assertion (c) is only stated for $r = 1$ and $r = 3$. Furthermore, the proof in [Mec09] seems to have some minor errors, which is why we provide a complete proof.

**Lemma 3.1.4.** Let $k \geq 1$ be an integer and let $h : \mathbb{R}^d \to \mathbb{R}$ be $k$-times differentiable with bounded derivatives $D^r h$ in any norm, $r = 1, \ldots, k$. Then, we have the following bounds for $r = 1, \ldots, k$:

(a) $M_r(f_h) \leq \frac{1}{r} M_r(h)$

(b) $\tilde{M}_2(f_h) \leq \frac{1}{2} \tilde{M}_2(h)$

If $\Sigma$ is additionally positive definite, then one even has for $r = 1, \ldots, k + 1$:

(c) $M_r(f_h) \leq \frac{2^{r-1} \Gamma(r/2)^2}{\sqrt{2\pi (r-1)!}} \|\Sigma^{-1/2}\|_{\text{op}} M_{r-1}(h)$

104
3.1. Stein’s method for multivariate normal distributions

(d) \( \tilde{M}_2(f_h) \leq \sqrt{\frac{2}{\pi}} \| \Sigma^{-1/2} \|_{\text{op}} M_1(h) \)

Remark 3.1.5. The bounds (c) and (d) show the smoothing properties of the operation \( h \mapsto f_h \) if \( \Sigma \) is regular. Then the solution \( f_h \) is of one order smoother than the test function \( h \).

Proof of Lemma 3.1.4. We use the notation from the proof of Lemma 3.1.3. Let \( x \in \mathbb{R}^d \) be arbitrarily given. By the assumed properties of \( h \), similarly as for (3.12) and (3.13), we have for each choice of \( i_1, \ldots, i_r \in \{1, \ldots, d\} \)

\[
\frac{\partial^r f_h}{\partial x_{i_r} \ldots \partial x_{i_1}}(x) = \frac{1}{2} \int_0^1 t^{\frac{r-1}{2}} E \left[ \frac{\partial^r h}{\partial x_{i_r} \ldots \partial x_{i_1}}(Z_{\Sigma,x,t}) \right] dt. \tag{3.15}
\]

Hence, for unit vectors \( u_1, \ldots, u_r \in \mathbb{R}^d \) we have

\[
D^r h(x)[u_1, \ldots, u_r] = \sum_{i_1, \ldots, i_r = 1}^d \frac{\partial^r f_h}{\partial x_{i_r} \ldots \partial x_{i_1}}(x)(u_1)_{i_1} \cdot \ldots \cdot (u_r)_{i_r} = \frac{1}{2} \int_0^1 t^{\frac{r-1}{2}} E \left[ \sum_{i_1, \ldots, i_r = 1}^d \frac{\partial^r h}{\partial x_{i_r} \ldots \partial x_{i_1}}(Z_{\Sigma,x,t})(u_1)_{i_1} \cdot \ldots \cdot (u_r)_{i_r} \right] dt
\]

and, hence,

\[
|D^r h(x)[u_1, \ldots, u_r]| \leq M_r(h) \frac{1}{2} \int_0^1 t^{r/2 - 1} dt.
\]

Since \( \int_0^1 t^{r/2 - 1} dt = \frac{2}{r} \), this immediately implies (a).

To prove (b), first note that by (3.15) we have

\[
\text{Hess} \ f_h(x) = \frac{1}{2} \int_0^1 E \left[ \text{Hess} \ h(Z_{\Sigma,x,t}) \right] dt.
\]

Hence, for any matrix \( A \in \mathbb{R}^{d \times d} \) we have

\[
|\langle \text{Hess} \ f_h(x), A \rangle_{\text{H.S.}}| = \frac{1}{2} \left| \int_0^1 E \left[ \langle \text{Hess} \ h(Z_{\Sigma,x,t}), A \rangle_{\text{H.S.}} \right] dt \right| \leq \frac{1}{2} \int_0^1 E \left[ |\langle \text{Hess} \ h(Z_{\Sigma,x,t}), A \rangle_{\text{H.S.}}| \right] dt \leq \frac{1}{2} \tilde{M}_2(h) \| A \|_{\text{H.S.}},
\]
3. Multivariate normal approximation and the classical compact groups

by the Cauchy-Schwarz inequality. Choosing \( A = \text{Hess} f_h(x) \) and dividing by \( \| \text{Hess} f_h(x) \|_{H.S} \) yields (b).

Next, we turn to the proof of (c). By (3.15) we have

\[
\frac{\partial^r f_h}{\partial x_{i_r} \ldots \partial x_{i_1}}(x) = \frac{1}{2} \int_0^1 t^{r-1} \int_{\mathbb{R}^d} \frac{\partial^r h}{\partial x_{i_r} \ldots \partial x_{i_1}}(\sqrt{t}x + \sqrt{1-t}Cy) \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}y^T y\right) dy dt.
\]

Let

\[
g := \frac{\partial^{r-1} h}{\partial x_{i_{r-1}} \ldots \partial x_{i_1}} \quad \text{and} \quad T(y) := \sqrt{t}x + \sqrt{1-t}Cy.
\]

Then, \( T \) is an affine \( C^1 \)-diffeomorphism with

\[
T^{-1}(z) = \frac{1}{\sqrt{1-t}}C^{-1}(z - \sqrt{t}x) \quad \text{and} \quad DT^{-1}(z) = \frac{1}{\sqrt{1-t}}C^{-1}.
\]

Let \( D := (d_{ij})_{i,j=1,\ldots,d} := C^{-1} = \Sigma^{-1/2} \). Then, by the chain rule formula

\[
\frac{\partial^r h}{\partial x_{i_r} \ldots \partial x_{i_1}}(\sqrt{t}x + \sqrt{1-t}Cy) = \frac{\partial g}{\partial x_{i_r}}(T(y)) = D(g \circ T)(y) \cdot \frac{1}{\sqrt{1-t}}C^{-1} \varepsilon_{i_r}
\]

\[
= \frac{1}{\sqrt{1-t}} \sum_{j=1}^d \frac{\partial (g \circ T)}{\partial y_j}(y) d_{j,i_r}
\]

\[
= \frac{1}{\sqrt{1-t}} \sum_{j=1}^d \frac{\partial}{\partial y_j} \left( \frac{\partial^{r-1} h}{\partial x_{i_{r-1}} \ldots \partial x_{i_1}}(\sqrt{t}x + \sqrt{1-t}Cy) \right) d_{j,i_r}.
\]

Plugging this into (3.16) and using the integration by parts formula as well as the symmetry of \( D \) yields
3.1. Stein’s method for multivariate normal distributions

\[
\frac{\partial^r f_h}{\partial x_{i_1} \ldots \partial x_{i_1}}(x) = \frac{1}{2} \int_0^1 \frac{t^{r-1}}{\sqrt{1 - t}} \int_{\mathbb{R}^d} \sum_{j=1}^d \frac{\partial}{\partial y_j} \left( \frac{\partial^{r-1} h}{\partial x_{i_{r-1}} \ldots \partial x_{i_1}}(\sqrt{tx + \sqrt{1 - t \text{C}}}) \right) \exp \left( -\frac{1}{2} y^T y \right) dy dt \]

\[
d_{j,i} \frac{1}{(2\pi)^{d/2}} \exp \left( -\frac{1}{2} y^T y \right) dy dt
\]

\[
= \frac{1}{2} \int_0^1 \frac{t^{r-1}}{\sqrt{1 - t}} \int_{\mathbb{R}^d} \sum_{j=1}^d \frac{\partial^{r-1} h}{\partial x_{i_{r-1}} \ldots \partial x_{i_1}}(\sqrt{tx + \sqrt{1 - t \text{C}}}) \left( \sum_{j=1}^d d_{i,j} y_j \right) \exp \left( -\frac{1}{2} y^T y \right) dy dt
\]

\[
\frac{1}{(2\pi)^{d/2}} \exp \left( -\frac{1}{2} y^T y \right) dy dt
\]

\[
= \frac{1}{2} \int_0^1 \frac{t^{r-1}}{\sqrt{1 - t}} \int_{\mathbb{R}^d} \sum_{j=1}^d \frac{\partial^{r-1} h}{\partial x_{i_{r-1}} \ldots \partial x_{i_1}}(\sqrt{tx + \sqrt{1 - t \text{C}}}) \left( D y \right)_{i_r} \exp \left( -\frac{1}{2} y^T y \right) dy dt
\]

\[
\frac{1}{(2\pi)^{d/2}} \exp \left( -\frac{1}{2} y^T y \right) dy dt
\]

By (3.17) for unit vectors \( u_1, \ldots, u_r \in \mathbb{R}^d \) we have

\[
D^r h(x)[u_1, \ldots, u_r] = \sum_{i_1, \ldots, i_r=1}^d \frac{\partial^r f_h}{\partial x_{i_1} \ldots \partial x_{i_1}}(x) (u_1)_{i_1} \cdot \ldots \cdot (u_r)_{i_r}
\]

\[
= \frac{1}{2} \int_0^1 \frac{t^{r-1}}{\sqrt{1 - t}} \sum_{i_1, \ldots, i_r=1}^d \frac{\partial^{r-1} h}{\partial x_{i_{r-1}} \ldots \partial x_{i_1}}(Z_{\Sigma, x, t})(u_1)_{i_1} \cdot \ldots \cdot (u_{r-1})_{i_{r-1}}
\]

\[
\sum_{i_r=1}^d (DZ)_{i_r} (u_r)_{i_r} \left( D Z, u_r \right) dt
\]

\[
= \frac{1}{2} \int_0^1 \frac{t^{r-1}}{\sqrt{1 - t}} \left[ D^{r-1} h(Z_{\Sigma, x, t})[u_1, \ldots, u_{r-1}] \cdot (DZ, u_r) \right] dt
\]
3. Multivariate normal approximation and the classical compact groups

and, hence,

$$|D^r h(x)[u_1, \ldots, u_r]| \leq \frac{1}{2} M_{r-1}(h) E\left[|\langle DZ, u_r \rangle|\right] \int_0^1 \frac{t^{\frac{r}{2} - 1}}{\sqrt{1 - t}} dt. \quad (3.18)$$

By the symmetry of $D$ we have

$$\langle DZ, u_r \rangle = (u^T_r D) Z = (Du_r)^T Z \sim N(0, \|Du_r\|_2^2)$$

and thus

$$E\left[|\langle DZ, u_r \rangle|\right] = \|Du_r\|_2 \sqrt{\frac{2}{\pi}} \leq \|D\|_{op} \sqrt{\frac{2}{\pi}} = \sqrt{\frac{2}{\pi}} \|\Sigma^{-1/2}\|_{op}. \quad (3.19)$$

Furthermore, by the relation between the Beta and Gamma functions $B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a + b)}$, the duplication formula $\Gamma(a) \Gamma(a + 1/2) = 2^{1-2a} \sqrt{\pi} \Gamma(2a)$ and from $\Gamma(1/2) = \sqrt{\pi}$ we have

$$\int_0^1 \frac{t^{\frac{r}{2} - 1}}{\sqrt{1 - t}} dt = \int_0^1 \frac{t^{\frac{r}{2} - 1}(1 - t)^{\frac{1}{2}}}{t^{\frac{r}{2}}} dt = B\left(\frac{r}{2}, \frac{1}{2}\right) = \frac{\Gamma\left(\frac{r}{2}\right) \Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{r + 1}{2}\right)} = \frac{\sqrt{\pi} \Gamma\left(\frac{r}{2}\right)^2}{2^{1-r} \sqrt{\pi} \Gamma(r)} = \frac{2^{r-1} \Gamma\left(\frac{r}{2}\right)}{(r - 1)!}. \quad (3.20)$$

From (3.18), (3.19) and (3.20) we can therefore conclude, that for each $x \in \mathbb{R}^d$ and for all unit vectors $u_1, \ldots, u_r \in \mathbb{R}^d$ it holds that

$$|D^r h(x)[u_1, \ldots, u_r]| \leq \frac{2^{r-1} \Gamma\left(\frac{r}{2}\right)}{\sqrt{2\pi(r - 1)!}} \|\Sigma^{-1/2}\|_{op} M_{r-1}(h).$$

This proves (c).

Finally, we prove part (d). From (3.17) we immediately obtain that for all $i, j \in \{1, \ldots, d\}$

$$\frac{\partial^2 f_h}{\partial x_i \partial x_j}(x) = \int_0^1 \frac{1}{2\sqrt{1 - t}} E \left[ \frac{\partial h}{\partial x_i}(Z_{\Sigma, x, t}) \cdot \left(\Sigma^{-1/2} Z\right)_j \right] dt$$

and, hence,

$$\text{Hess } f_h(x) = \int_0^1 \frac{1}{2\sqrt{1 - t}} E \left[ \nabla h(Z_{\Sigma, x, t}) \left(\Sigma^{-1/2} Z\right)^T \right] dt. \quad (3.21)$$
Thus, for any matrix $A \in \mathbb{R}^{d \times d}$ we obtain

\[
\langle \text{Hess} f_h(x), A \rangle_{\text{H.S.}} = \int_0^1 \frac{1}{2\sqrt{1-t}} E \left[ \text{Tr} \left( \nabla h(Z_{\Sigma,x,t}) \left( \Sigma^{-1/2} Z \right)^T A^T \right) \right] dt
\]

\[
= \int_0^1 \frac{1}{2\sqrt{1-t}} E \left[ \text{Tr} \left( \nabla h(Z_{\Sigma,x,t}) \left( A \Sigma^{-1/2} Z \right)^T \right) \right] dt
\]

\[
= \int_0^1 \frac{1}{2\sqrt{1-t}} E \left[ \left( A \Sigma^{-1/2} Z \right)^T \nabla h(Z_{\Sigma,x,t}) \right] dt
\]

\[
= \int_0^1 \frac{1}{2\sqrt{1-t}} E \left[ \langle A \Sigma^{-1/2} Z, \nabla h(Z_{\Sigma,x,t}) \rangle \right] dt
\] (3.22)

and

\[
\left| \langle \text{Hess} f_h(x), A \rangle_{\text{H.S.}} \right| \leq M_1(h) E \left[ \| A \Sigma^{-1/2} Z \|_2 \right] \int_0^1 \frac{1}{2\sqrt{1-t}} dt
\]

\[
= M_1(h) E \left[ \| A \Sigma^{-1/2} Z \|_2 \right]. \quad (3.23)
\]

Let $B = (b_{ij})_{i,j=1,...,d} := A \Sigma^{-1/2}$. Then, by Jensen’s inequality

\[
E \left[ \| BZ \|_2 \right] = E \left[ \left( \langle BZ, BZ \rangle \right)^{1/2} \right] \leq \left( E \left[ \langle BZ, BZ \rangle \right] \right)^{1/2}
\]

\[
= \left( \sum_{i=1}^d \left( \sum_{j=1}^d b_{ij} Z_j \right)^2 \right)^{1/2} = \left( \sum_{i,j=1}^d b_{ij} b_{il} E[Z_j Z_l] \right)^{1/2}
\]

\[
= \left( \sum_{i,j=1}^d b_{ij}^2 \right)^{1/2} = \| B \|_{\text{H.S.}}. \quad (3.24)
\]

Now, note that for matrices $A, B \in \mathbb{R}^{d \times d}$ one has the inequality

\[
\| AB \|_{\text{H.S.}} \leq \| A \|_{\text{H.S.}} \| B \|_{\text{op}}. \quad (3.25)
\]

From (3.23), (3.24) and (3.25) we thus obtain

\[
\left| \langle \text{Hess} f_h(x), A \rangle_{\text{H.S.}} \right| \leq M_1(h) \| A \Sigma^{-1/2} \|_{\text{H.S.}} \leq M_1(h) \| A \|_{\text{H.S.}} \| \Sigma^{-1/2} \|_{\text{op}}.
\]

Again, choosing $A := \text{Hess} f_h(x)$ and dividing by $\| \text{Hess} f_h(x) \|_{\text{H.S.}}$ yield the desired result.
3. Multivariate normal approximation and the classical compact groups

If one prefers using the quantities \(|h|_r|\) in one’s bounds, the following lemma provides useful bounds. For example, one may reduce the number of existing derivatives of the test function \(h\) from 3 to 2 in Theorem 2.1 from [RR09], if \(\Sigma\) is positive definite and \(\|\Sigma^{-1/2}\|_{\text{op}}\) may be computed or, at least, be sufficiently bounded.

**Lemma 3.1.6.** Let \(k \geq 1\) be an integer and let \(h : \mathbb{R}^d \to \mathbb{R}\) be \(k\)-times differentiable with bounded derivatives \(D^rh\) for each \(r = 1, \ldots, k\).

(a) For each \(r \in \{1, \ldots, k\}\) we have \(|f_h|_r \leq \frac{1}{r} |h|_r|\).

(b) If \(\Sigma\) is additionally positive definite, then for all \(r = 1, \ldots, k+1\) it holds that

\[
|f_h|_r \leq \frac{2^{r-1} \Gamma\left(\frac{r}{2}\right)^2}{\sqrt{2\pi(r-1)!}} \|\Sigma^{-1/2}\|_{\text{op}} |h|_{r-1}.
\]

**Proof.** Part (a) immediately follows from (3.15) since \(\frac{1}{2} \int_0^1 t^{r/2-1} dt = \frac{1}{r}\).

To prove (b), let \(1 \leq i_1, \ldots, i_r \leq d\) be given. By (3.17) with \(D = \Sigma^{-1/2}\), we have for each \(x \in \mathbb{R}^d\)

\[
\frac{\partial^r f_h}{\partial x_{i_r} \ldots \partial x_{i_1}}(x) = \frac{1}{2} \int_0^1 \frac{t^{r-1}}{\sqrt{1-t}} E \left[ \frac{\partial^{r-1} h}{\partial x_{i_{r-1}} \ldots \partial x_{i_1}}(Z_{x,t}) \cdot (DZ)_{i_r} \right] dt
\]

and, hence, using (3.20)

\[
\left\| \frac{\partial^r f_h}{\partial x_{i_r} \ldots \partial x_{i_1}} \right\|_{\infty} \leq \frac{1}{2} \int_0^1 \frac{t^{r-1}}{\sqrt{1-t}} dt \left\| \frac{\partial^{r-1} h}{\partial x_{i_{r-1}} \ldots \partial x_{i_1}} \right\|_{\infty} E\left[ |(DZ)_{i_r}| \right] 
\leq \frac{1}{2} \int_0^1 \frac{t^{r-1}}{\sqrt{1-t}} dt |h|_{r-1} E\left[ |(DZ)_{i_r}| \right] 
= \frac{2^{r-2} \Gamma\left(\frac{r}{2}\right)^2}{(r-1)!} |h|_{r-1} E\left[ |(DZ)_{i_r}| \right] 
= 2^{r-2} \Gamma\left(\frac{r}{2}\right)^2 |h|_{r-1} E\left[ |(DZ)_{i_r}| \right] 
(3.26)
\]

Now, let the \(i_r\)-th row vector of the matrix \(D\) be \(a^T = (a_1, \ldots, a_d)\). Then,

\[(DZ)_{i_r} = a^T Z = \sum_{j=1}^d a_j Z_j \sim N(0, \|a\|_2^2)\]

and, thus

110
3.1. Stein’s method for multivariate normal distributions

\[ E[(DZ)_r] = \|a\|_2 \sqrt{\frac{2}{\pi}} \leq \|D\|_{\text{op}} \sqrt{\frac{2}{\pi}}, \]  \hspace{1cm} (3.27)

since for every matrix \( D \), the operator norm is not less than the Euclidian norm of any of its rows. The claim now follows from (3.26) and (3.27).

3.1.2. The exchangeable pairs approach for multivariate normal approximation

In the following we present the method of exchangeable pairs in the context of multivariate normal approximation. Although this coupling technique lies at the heart of univariate normal approximation by Stein’s method (see Stein’s monograph [Ste86]), it was only in 2008 by Chatterjee and Meckes [CM08] that the problem of developing a suitable technique in the multivariate case was finally attacked. In their work, for a given random vector \( W = (W_1, \ldots, W_d)^T \) they assume the existence of another random vector \( W' = (W'_1, \ldots, W'_d) \), defined on the same probability space, such that \( W' \) has the same distribution as \( W \) and such that the following linear regression property

\[ E[W' - W | W] = -\lambda W \]  \hspace{1cm} (3.28)

is satisfied for some positive constant \( \lambda \). Under these assumptions the authors prove various so-called plug-in theorems to bound the distance from \( W \) to a standard normal random vector.

In [RR09] Reinert and Röllin motivated and investigated the more general linear regression property

\[ E[W' - W | W] = -\Lambda W + R, \]  \hspace{1cm} (3.29)

where now \( \Lambda \) is an invertible non-random \( d \times d \) matrix and \( R = (R_1, \ldots, R_d)^T \) is a small remainder term. However, in contrast to Chatterjee and Meckes, Reinert and Röllin need the full strength of the exchangeability of the vector \((W, W')\). Finally, in [Mec09], Elizabeth Meckes reconciled the two approaches, allowing for the more general linear regression property from [RR09] and using sharper coordinate-free bounds relying on the bounds on the solution to the Stein equation given by Lemma 3.1.4. In what follows, we will present the theory from Meckes’ paper [Mec09] and will specifically formulate the corresponding plug-in theorems. It should be noted, that both papers, [CM08] and [Mec09], contain an infinitesimal version of Stein’s method of exchangeable pairs, which can be applied if for each \( t > 0 \) an exchangeable pair \((W, W_t)\) of random vectors is given, such that a suitable
3. Multivariate normal approximation and the classical compact groups

infinitesimal linear regression property and further limiting conditions hold true. The idea of this infinitesimal version goes back to Stein [Ste95] and was further developed for univariate normal approximation by Meckes in her PhD thesis [Mec06].

Recall, that for two probability measures $\mu$ and $\nu$ on $(\mathbb{R}^d, \mathcal{B}^d)$, whose expectations both exist, one defines their Wasserstein distance by

$$d_W(\mu, \nu) := \sup \left\{ \left| \int_{\mathbb{R}^d} g \, d\mu - \int_{\mathbb{R}^d} g \, d\nu \right| : M_1(g) \leq 1 \right\}$$

(3.30)

and for random vectors $X$ and $Y$ with existing expectations one lets

$$d_W(X, Y) := d_W(\mathcal{L}(X), \mathcal{L}(Y)) = \sup \left\{ \left| E[g(X)] - E[g(Y)] \right| : M_1(g) \leq 1 \right\}.$$

It is well-known that on the space of probability measures with existing expectations, the topology induced by the Wasserstein distance is stronger than the topology induced by weak convergence of probability measures (see, e.g. [Dud02]). For random matrices $M_n, M$ ($n \in \mathbb{N}$) defined on the same probability space $(\Omega, \mathcal{A}, P)$ we will say that $(M_n)_{n \in \mathbb{N}}$ converges to $M$ in $L^1(\|\cdot\|_{\text{H.S.}})$ if $\|M_n - M\|_{\text{H.S.}}$ converges to 0 in $L^1(P)$ as $n \to \infty$.

The following two plug-in theorems are (versions of) Theorems 3 and 4 of [Mec09]. We also refer there for the proofs of these assertions.

**Theorem 3.1.7.** Let $(W, W')$ be an exchangeable pair of $\mathbb{R}^d$-valued $L^2(P)$ random vectors defined on a probability space $(\Omega, \mathcal{A}, P)$ and let $\mathcal{F} \subseteq \mathcal{A}$ be a sub-$\sigma$-algebra of $\mathcal{A}$ such that $\sigma(W) \subseteq \mathcal{F}$. Suppose there is a non-random invertible matrix $\Lambda \in \mathbb{R}^{d \times d}$, a non-random positive semidefinite matrix $\Sigma$, an $\mathcal{F}$-measurable random vector $R$ and an $\mathcal{F}$-measurable random matrix $S$ such that (3.29) and

$$E\left[ (W' - W)(W' - W)^T \mid \mathcal{F} \right] = 2\Lambda \Sigma + S$$

(3.31)

hold true.

(a) For any $h \in C^3(\mathbb{R}^d)$ such that $E[|h(W)|] < \infty$ and $E[|h(Z_\Sigma)|] < \infty$,
3.1. Stein’s method for multivariate normal distributions

\[ |E[h(W)] - E[h(Z_\Sigma)]| \leq \|\Lambda^{-1}\|_{\text{op}} \left( M_1(h)E[\|R\|_2] + \frac{1}{4}M_2(h)E[\|S\|_{\text{H.S.}}] \right. \\
+ \frac{1}{18}M_3(h)E[\|W' - W\|_2^3] \bigg) \\
\leq \|\Lambda^{-1}\|_{\text{op}} \left( M_1(h)E[\|R\|_2] + \frac{\sqrt{d}}{4}M_2(h)E[\|S\|_{\text{H.S.}}] \right. \\
+ \frac{1}{18}M_3(h)E[\|W' - W\|_2^3] \bigg).

(b) If \( \Sigma \) is actually positive definite, then for each \( h \in C^2(\mathbb{R}^d) \) such that 
\( E[|h(W)|] < \infty \) and \( E[|h(Z_\Sigma)|] < \infty \),

\[ |E[h(W)] - E[h(Z_\Sigma)]| \leq M_1(h)\|\Lambda^{-1}\|_{\text{op}} \left( E[\|R\|_2] + \frac{\|\Sigma^{-1/2}\|_{\text{op}}}{\sqrt{2\pi}}E[\|S\|_{\text{H.S.}}] \right) \\
+ \frac{\sqrt{2\pi}}{24}M_2(h)\|\Lambda^{-1}\|_{\text{op}}\|\Sigma^{-1/2}\|_{\text{op}}E[\|W' - W\|_2^3] \]
3. Multivariate normal approximation and the classical compact groups

Semidefinite matrix $\Sigma$, a random vector $R = (R_1, \ldots, R_d)^T$, a random $d \times d$-matrix $S$, a sub-$\sigma$-algebra $\mathcal{F}$ of $\mathcal{A}$ with $\sigma(W) \subseteq \mathcal{F}$ and a deterministic function $s : (0, \infty) \to (0, \infty)$ such that the following three conditions are satisfied:

\[
\frac{1}{s(t)} E[W_t - W|\mathcal{F}] \xrightarrow{t \to 0} -\Lambda W + R \text{ in } L^1(P). \tag{i}
\]

\[
\frac{1}{s(t)} E[(W_t - W)(W_t - W)^T |\mathcal{F}] \xrightarrow{t \to 0} 2\Lambda \Sigma + S \text{ in } L^1(\|\cdot\|_{\text{H.S.}}). \tag{ii}
\]

\[
\lim_{t \to 0} \frac{1}{s(t)} E\left[\|W_t - W\|_2^2 1_{\{\|W_t - W\|_2 > \varepsilon\}}\right] = 0 \text{ for each } \varepsilon > 0. \tag{iii}
\]

Then:

(a) For each $h \in C^2(\mathbb{R}^d)$ such that $E[|h(W)|] < \infty$ and $E[|h(Z_\Sigma)|] < \infty$,

\[
\left|E[h(W)] - E[h(Z_\Sigma)]\right| \\
\leq \|\Lambda^{-1}\|_{\text{op}} \left(M_1(h) E[\|R\|_2] + \frac{1}{4} M_2(h) E[\|S\|_{\text{H.S.}}]\right) \\
\leq \|\Lambda^{-1}\|_{\text{op}} \left(M_1(h) E[\|R\|_2] + \frac{\sqrt{d}}{4} M_2(h) E[\|S\|_{\text{H.S.}}]\right).
\]

(b) If $\Sigma$ is actually positive definite, then

\[
d_{W}(W, Z_\Sigma) \leq \|\Lambda^{-1}\|_{\text{op}} \left(E[\|R\|_2] + \frac{1}{\sqrt{2\pi}} \|\Sigma^{-1/2}\|_{\text{op}} E[\|S\|_{\text{H.S.}}]\right).
\]

Remark 3.1.10. When applying Theorem 3.1.9 it is often easier to verify the following stronger condition in the place of Condition (iii):

\[
\lim_{t \to 0} \frac{1}{s(t)} E\left[\|W_t - W\|_2^2\right] = 0. \tag{iii}'
\]

For our applications we will need a version of Theorem 3.1.9 for random vectors with values in $\mathbb{C}^d$. To this end, we will introduce the general multivariate complex normal distribution. First, however, we have to fix some notation and review some facts from linear algebra.

For a vector $z = (z_1, \ldots, z_d)^T \in \mathbb{C}^d$ with $z_j = x_j + iy_j$ and $x_j, y_j \in \mathbb{R}$, $j = 1, \ldots, d$, we denote by
3.1. Stein’s method for multivariate normal distributions

\[ z_R := (x_1, y_1, \ldots, x_d, y_d)^T := (\text{Re}(z_1), \text{Im}(z_1), \ldots, \text{Re}(z_d), \text{Im}(z_d))^T \in \mathbb{R}^{2d} \]

the real version of \( z \). This gives rise to an isomorphism \( \Phi : \mathbb{C}^d \rightarrow \mathbb{R}^{2d} \) of real vector spaces, given by \( \Phi(z) := z_R \). If \( A = (a_{jk})_{j,k=1,\ldots,d} \in \mathbb{C}^{d \times d} \) is a matrix, then the corresponding \( \mathbb{C} \)-linear mapping \( \varphi_A : \mathbb{C}^d \rightarrow \mathbb{C}^d \) with \( \varphi_A(z) := Az \) is also \( \mathbb{R} \)-linear and, hence, is the composition \( \psi := \Phi \circ \varphi_A \circ \Phi^{-1} : \mathbb{R}^{2d} \rightarrow \mathbb{R}^{2d} \). Thus, there is a unique matrix \( B \in \mathbb{R}^{2d \times 2d} \) such that \( \psi(x) = Bx \) for each \( x \in \mathbb{R}^{2d} \). If we write \( A_R := A \), then this simply means that

\[ (Az)_R = A_R z_R \quad \text{for each} \quad z \in \mathbb{C}^d. \]

It should be noted that in the case \( d = 1 \) using our notation, it has to be specified if a complex number is interpreted as a matrix (a linear transformation) or as a vector, since the corresponding real object would be a \( 2 \times 2 \)-matrix in the first case, but a vector in \( \mathbb{R}^2 \) in the latter case. In fact, it can easily be checked, that \( A_R \) is a matrix of \( d^2 \) \( 2 \times 2 \)-blocks with the block at position \((j, k)\) given by

\[
\begin{pmatrix}
\text{Re}(a_{jk}) & -\text{Im}(a_{jk}) \\
\text{Im}(a_{jk}) & \text{Re}(a_{jk})
\end{pmatrix},
\]

(3.32)

We will have to extend the matrix norms \( \|\cdot\|_{\text{op}} \) and \( \|\cdot\|_{\text{H.S.}} \) to \( \mathbb{C}^{d \times d} \). For matrices \( A, B \in \mathbb{C}^{d \times d} \) we let

\[ \|A\|_{\text{op}} := \sup \left\{ \|Az\|_2 : z \in \mathbb{C}^d \text{ and } \|z\|_2 = 1 \right\} \]

and

\[ \langle A, B \rangle_{\text{H.S.}} := \text{Tr}(AB^*) = \text{Tr}(B^*A) = \text{Tr}(BA^*) = \sum_{j,k=1}^{d} a_{jk}b_{jk}. \]

Here and in what follows, we let \( A^* := A^T \) denote the Hermitian adjoint of the matrix \( A \). Thus, \( \langle \cdot, \cdot \rangle_{\text{H.S.}} \) is just the standard inner product on the space \( \mathbb{C}^{d \times d} \cong \mathbb{C}^d \). Again, we denote by \( \|\cdot\|_{\text{H.S.}} \) the corresponding norm.

The following lemma provides several rules for the operation \( A \mapsto A_R \), which will be needed in the sequel.

**Lemma 3.1.11.** Let \( A, B \in \mathbb{C}^{d \times d} \).

1. \( (AB)_R = A_RB_R \)
2. If \( A \) is invertible, then so is \( A_R \) and \( (A_R)^{-1} = (A^{-1})_R \).
3. Multivariate normal approximation and the classical compact groups

(3) \((A^*)_R = (A_R)^T\)

(4) \((AA^*)_R = A_R(A^*_R) = A_R(A_R)^T\)

(5) If \(A\) is Hermitian, then \(A_R\) is symmetric.

(6) If \(A\) is positive (semi-)definite, then so is \(A_R\).

(7) \(\|A_R\|_{\text{op}} = \|A\|_{\text{op}}\)

(8) \(\|A_R\|_{\text{H.S.}} = \sqrt{2}\|A\|_{\text{H.S.}}\).

Proof. To prove (1) let \(z \in \mathbb{C}^d\) be given. Then, by associativity of matrix multiplication, we have

\[(A_R B_R)z_R = A_R (B_R z_R) = A_R (Bz)_R = (A (Bz))_R = ((AB)z)_R = (AB)_R z_R.\]

Hence, by uniqueness we have \((AB)_R = (A_R B_R)\) which proves (1).

To prove (2) let \(C := A^{-1}\). Then, using (1)

\[I_{2d} = (I_d)_R = (AC)_R = A_R C_R,\]

which implies that \(A_R\) is invertible and \((A_R)^{-1} = C_R = (A^{-1})_R\).

As to (3) note, that by (3.32) the block at position \((j, k)\) in \((A^*)_R\) is given by

\[\begin{pmatrix}
\text{Re}(a_{kj}) & -\text{Im}(a_{kj}) \\
\text{Im}(a_{kj}) & \text{Re}(a_{kj})
\end{pmatrix} = \begin{pmatrix}
\text{Re}(a_{kj}) & \text{Im}(a_{kj}) \\
-\text{Im}(a_{kj}) & \text{Re}(a_{kj})
\end{pmatrix}\]

This is the transpose of the block at position \((k, j)\) of \(A_R\), again by (3.32). Hence, (3) follows by transposition rules for block matrices.

Next, (4) follows immediately from (3) and (1) and (5) follows from (3).

To prove (6), first let \(A\) be positive semidefinite. Then, there exists a unique Hermitian square root \(C \in \mathbb{C}^{d \times d}\) to \(A\), i.e. \(C^2 = CC^* = A\). Hence, by (4) we have

\[C_R (C_R)^T = (CC^*)_R = A_R.\]

This implies that \(A_R\) is symmetric and in fact positive semidefinite. If \(A\) is even positive definite, then by the just proven, \(A_R\) is positive semidefinite and is additionally invertible by (2). Hence, also \(A_R\) is positive definite.
3.1. Stein’s method for multivariate normal distributions

The claim of (7) follows from \( \|z\|_2 = \|z_R\|_2 \) for each \( z \in \mathbb{C}^d \).
As to (8) note that by the block structure of \( A_R \) given by (3.32) we have

\[
\|A_R\|_{\text{H.S.}}^2 = \sum_{j,k=1}^d \left( 2(\text{Re}(a_{jk}))^2 + 2(\text{Im}(a_{jk}))^2 \right) = 2 \sum_{j,k=1}^d \left( (\text{Re}(a_{jk}))^2 + (\text{Im}(a_{jk}))^2 \right)
= 2 \sum_{j,k=1}^d |a_{jk}|^2 = 2\|A\|_{\text{H.S.}}^2,
\]
proving the claim.

Now, also let \( w = (w_1, \ldots, w_d)^T \in \mathbb{C}^d \) be given with \( w_j = u_j + iv_j \) and \( u_j, v_j \in \mathbb{R} \) for \( j = 1, \ldots, d \). Then, by the rules of matrix multiplication we obtain that the real \( 2d \times 2d \)-matrix \( z_R w_R^T \) is a matrix of \( d \times 2 \times 2 \)-blocks with the block at position \((j, k)\) given by

\[
\begin{pmatrix}
  x_j u_k & x_j v_k \\
y_j u_k & y_j v_k
\end{pmatrix} = \frac{1}{2}
\begin{pmatrix}
  \text{Re}(z_j \overline{w}_k + z_j w_k) & \text{Im}(z_j w_k - z_j \overline{w}_k) \\
  \text{Im}(z_j w_k + z_j \overline{w}_k) & \text{Re}(z_j \overline{w}_k - z_j w_k)
\end{pmatrix}
= \frac{1}{2}
\begin{pmatrix}
  \text{Re}(z_j w_k) - \text{Im}(z_j \overline{w}_k) & \text{Im}(z_j w_k) - \text{Re}(z_j \overline{w}_k) \\
  \text{Im}(z_j \overline{w}_k) & \text{Re}(z_j w_k)
\end{pmatrix}
+ \frac{1}{2}
\begin{pmatrix}
  \text{Re}(z_j w_k) + \text{Im}(z_j \overline{w}_k) & \text{Im}(z_j w_k) + \text{Re}(z_j \overline{w}_k) \\
  \text{Im}(z_j \overline{w}_k) & \text{Re}(z_j w_k)
\end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
\]

Thus, by block multiplication we have

\[
z_R w_R^T = \frac{1}{2} (zw^*)_R + \frac{1}{2} (zw^T)^*_R \cdot J,
\]

where \( J \) is the Kronecker product

\[
J := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes I_d \in \mathbb{R}^{2d \times 2d}.
\]

Now, we return to random vectors. First, let \( X_1, Y_1, \ldots, X_d, Y_d \) be independent, identically distributed, real-valued random variables with distribution \( N(0, 1/2) \). Then, with \( Z_j := X_j + iY_j, j = 1, \ldots, d \), the vector \( Z := (Z_1, \ldots, Z_d)^T \) has the \( d \)-dimensional complex standard normal distribution, in symbols

\[
Z \sim N_{d, \mathbb{C}}(0, I_d).
\]

Equivalently, the random vector \( Z_R \in \mathbb{R}^{2d} \) has distribution \( N_{2d}(0, \frac{1}{2} I_d) \). It easily follows from the definition that for \( Z \sim N_{d, \mathbb{C}}(0, I_d) \) one has \( E[Z] = 0 \),

\[
117
\]
We say, that a random vector \( Y = (Y_1, \ldots, Y_d)^T \) with values in \( \mathbb{C}^d \) has a \( d \)-dimensional complex normal distribution, if there exist a matrix \( A \in \mathbb{C}^{d \times d} \), a vector \( b \in \mathbb{C}^d \) and a random vector \( Z \sim N_{d,\mathbb{C}}(0, I_d) \) such that \( Y = AZ + b \). Note that for such a vector \( Y \) we have \( E[Y] = AE[Z] + b = b \) and furthermore its covariance matrix \( \Sigma_Y \) is given by

\[
\Sigma_Y := \text{Cov}(Y) = E[(Y - b)(Y - b)^*] = E[(AZ)(AZ)^*] = AE[ZZ^*]A^* = AA^*
\]

by (3.34). Hence, \( \Sigma_Y \) is always Hermitian positive semidefinite.

**Proposition 3.1.12.** Let a vector \( b \in \mathbb{C}^d \) and a positive semidefinite matrix \( \Sigma \in \mathbb{C}^{d \times d} \) be given. Then, there exists a unique \( d \)-dimensional complex normal distribution on a random vector \( Y \) with values in \( \mathbb{C}^d \) such that \( E[Y] = b \) and \( \Sigma_Y = \Sigma \). This distribution will be denoted by \( N_{d,\mathbb{C}}(b, \Sigma) \).

**Proof.** To prove the existence, let \( A \) be the unique Hermitian square root of \( \Sigma \) and take \( Z \sim N_{d,\mathbb{C}}(0, I_d) \). Then, with \( Y := AZ + b \) by (3.36) we have \( E[Y] = b \) and

\[
\Sigma_Y = AA^* = A^2 = \Sigma.
\]

Thus, the vector \( Y \) has the desired properties.

Conversely, if \( Y \) is such a vector, then there exist a matrix \( A \in \mathbb{C}^{d \times d} \) and a random vector \( Z \sim N_{d,\mathbb{C}}(0, I_d) \) with \( Y = AZ + b \) and \( AA^* = \Sigma \). Thus,

\[
Y_R = \begin{pmatrix} AZ + b \end{pmatrix}_R = A_RZ_R + b_R = \frac{1}{\sqrt{2}}A_R\bar{Z} + b_R,
\]

where \( \bar{Z} := \sqrt{2}Z_R \sim N_{2d}(0, I_{2d}) \). Hence, \( Y_R \) has a \( 2d \)-dimensional real normal distribution with expected value \( E[Y_R] = E[Y]_R = b_R \) and covariance matrix

\[
C := \left( \frac{1}{\sqrt{2}}A_R \right) \left( \frac{1}{\sqrt{2}}A_R^T \right) = \frac{1}{2}A_R(A_R^T) = \frac{1}{2}(AA^*)_R = \frac{1}{2}\Sigma_R.
\]

By the uniqueness result for real multivariate normal distributions we know that the distribution of \( Y_R \) is determined by \( b_R \) and \( \frac{1}{2}\Sigma_R \) and hence, also by \( b \) and \( \Sigma \). This clearly implies that also the distribution of \( Y = \Phi^{-1}(Y_R) \) is determined by \( b \) and \( \Sigma \).
3.1. Stein’s method for multivariate normal distributions

For an $L^2(P)$ random vector $Y$ with values in $\mathbb{C}^d$ (not necessarily complex normally distributed) let

$$\Gamma_Y := E[(Y - E[Y])(Y - E[Y])^T] \quad \text{and} \quad \Sigma_Y := \text{Cov}(Y) = E[(Y - E[Y])(Y - E[Y])^*].$$

If $Y = AZ + b$ is normally distributed, then by (3.35) and (3.36)

$$\Gamma_Y = E[(AZ)(AZ)^T] = AE[ZZ^T]A^T = 0 \quad \text{and} \quad \Sigma_Y = AA^*.$$ (3.37)

**Proposition 3.1.13.** Let $Y$ be a random vector with values in $\mathbb{C}^d$ such that the real version $Y_\mathbb{R}$ has a 2$\times$ dimensional multivariate normal distribution $N_{2d}(\mu, C)$ for a vector $\mu \in \mathbb{R}^{2d}$ and a positive semidefinite matrix $C \in \mathbb{R}^{2d \times 2d}$. Then $Y$ has a multivariate complex normal distribution if and only if $\Gamma_Y = 0$. In this case we have $C = \frac{1}{2}(\Sigma_Y)_{\mathbb{R}}$.

**Proof.** If $Y$ is a complex normal random vector, then we know from (3.37) that $\Gamma_Y = 0$.

Conversely, let $\Gamma_Y = 0$. Then using (3.33) we have

\[
C = \text{Cov}(Y_\mathbb{R}) = E[(Y_\mathbb{R} - \mu)(Y_\mathbb{R} - \mu)^T] \\
= E[(Y_\mathbb{R} - E[Y_\mathbb{R}](Y_\mathbb{R} - E[Y_\mathbb{R}])^T] \\
= \frac{1}{2}E[(Y - E[Y])(Y - E[Y])^*]_{\mathbb{R}} + \frac{1}{2}E[(Y - E[Y])(Y - E[Y])^T]_{\mathbb{R}} \cdot J \\
= \frac{1}{2}(\Sigma_Y)_{\mathbb{R}} + \frac{1}{2}(\Gamma_Y)_{\mathbb{R}} \cdot J \\
= \frac{1}{2}(\Sigma_Y)_{\mathbb{R}}.
\]

Since $Y_\mathbb{R}$ has distribution $N_{2d}(\mu, C)$, there exists a vector 

$$\tilde{Z} = (X_1, Y_1, \ldots, X_d, Y_d)^T \sim N_{2d}(0, I_{2d})$$

such that

$$Y_\mathbb{R} = C^{1/2} \tilde{Z} + \mu = \frac{1}{\sqrt{2}} \left( \Sigma_Y \right)^{1/2}_{\mathbb{R}} \tilde{Z} + \mu.$$

Now choose $b \in \mathbb{C}^d$ with $b_\mathbb{R} = \mu$ and let $Z := \frac{1}{\sqrt{2}}(X_1 + iY_1, \ldots, X_d + iY_d)^T$. Then, $Z \sim N_{d,\mathbb{C}}(0, I_d)$ and

$$Y_\mathbb{R} = \frac{1}{\sqrt{2}} \left( \Sigma_Y \right)^{1/2}_{\mathbb{R}} \tilde{Z} + \mu = \frac{1}{\sqrt{2}} \left( \Sigma_Y \right)^{1/2}_{\mathbb{R}} \sqrt{2}Z_\mathbb{R} + b_\mathbb{R} = \left( \Sigma_Y^{1/2}Z + b \right)_{\mathbb{R}}$$

119
and, hence, $Y = \Sigma Y + b$ is a multivariate complex normal random vector.

\begin{corollary}
Let the random vector $Y$ have the $d$-dimensional complex normal distribution $N_{d, \mathbb{C}}(b, \Sigma)$ for some $b \in \mathbb{C}^d$ and some positive semidefinite matrix $\Sigma \in \mathbb{C}^{d \times d}$. Let the real version of $Y$ be given by $Y_\mathbb{R} = (X_1, X_2, \ldots, X_{2d-1}, X_{2d})^T$. Then, $X_{2j-1}$ and $X_{2j}$ are uncorrelated and, hence, independent for each $1 \leq j \leq d$.
\end{corollary}

\begin{proof}
By Proposition 3.1.13 the covariance matrix of $Y_\mathbb{R}$ is $C := (c_{jk})_{j,k=1,\ldots,2d} = \frac{1}{2}(\Sigma Y)_\mathbb{R}$. Let $j \in \{1, \ldots, d\}$ be given. Since $C$ is symmetric, we must have $c_{2j-1,2j} = c_{2j,2j-1}$. But from (3.32) we know that

$$c_{2j-1,2j} = -\text{Im}(\sigma_{jj}) = -c_{2j,2j-1},$$

where $\Sigma = (\sigma_{j,k})_{j,k=1,\ldots,d}$. Thus,

$$\text{Cov}(X_{2j-1}, X_j) = c_{2j-1,2j} = 0.$$

\end{proof}

Now we are in the position to state and prove versions of Theorems 3.1.7 and 3.1.9 for random vectors with values in $\mathbb{C}^d$. In what follows, we let $Z \sim N_{d, \mathbb{C}}(0, I_d)$ and for a positive semidefinite matrix $\Sigma \in \mathbb{C}^{d \times d}$ we let $Z_\Sigma := \Sigma^{1/2}Z$. Furthermore, for a function $h : \mathbb{R}^{2d} \to \mathbb{R}$ and $z \in \mathbb{C}^d$ we write $h(z)$ for $h(z_\mathbb{R})$.

\begin{theorem}
Let $(W, W')$ be an exchangeable pair of $\mathbb{C}^d$-valued $L^2(P)$ random vectors defined on a probability space $(\Omega, \mathcal{A}, P)$ and let $\mathcal{F} \subseteq \mathcal{A}$ be a sub-$\sigma$-algebra of $\mathcal{A}$ such that $\sigma(W) \subseteq \mathcal{F}$. Suppose there are a non-random invertible matrix $\Lambda \in \mathbb{C}^{d \times d}$, a non-random positive semidefinite matrix $\Sigma \in \mathbb{C}^{d \times d}$, an $\mathcal{F}$-measurable random vector $R$ and $\mathcal{F}$-measurable random matrices $S$ and $T$ such that the following properties hold:

\begin{align*}
E\left[W' - W \mid \mathcal{F}\right] &= -\Lambda W + R \tag{i} \\
E\left[(W' - W)(W' - W)^* \mid \mathcal{F}\right] &= 2\Lambda \Sigma + S \tag{ii} \\
E\left[(W' - W)(W' - W)^T \mid \mathcal{F}\right] &= T \tag{iii}
\end{align*}

(a) For any $h \in C^3(\mathbb{R}^{2d})$ such that $E[|h(W)|] < \infty$ and $E[|h(Z_\Sigma)|] < \infty$, 

\end{theorem}
3.1. Stein’s method for multivariate normal distributions

\[ |E[h(W)] - E[h(Z)]| \]
\[ \leq \|\Lambda^{-1}\|_{op} \left( M_1(h)E[\|R\|_2] + \frac{1}{4\sqrt{2}} M_2(h)E[\|S\|_{H.S.} + \|T\|_{H.S.}] \right. \]
\[ \left. + \frac{1}{18} M_3(h)E[\|W' - W\|_2^3] \right) \]
\[ \leq \|\Lambda^{-1}\|_{op} \left( M_1(h)E[\|R\|_2] + \frac{\sqrt{d}}{4} M_2(h)E[\|S\|_{H.S.} + \|T\|_{H.S.}] \right. \]
\[ \left. + \frac{1}{18} M_3(h)E[\|W' - W\|_2^3] \right). \]

(b) If \( \Sigma \) is actually positive definite, then for each \( h \in C^2(\mathbb{R}^{2d}) \) such that \( E[|h(W)|] < \infty \) and \( E[|h(Z)|] < \infty \),

\[ |E[h(W)] - E[h(Z)]| \leq M_1(h)\|\Lambda^{-1}\|_{op} \left( E[\|R\|_2] \right. \]
\[ \left. + \frac{\|\Sigma^{-1/2}\|_{op} E[\|S\|_{H.S.} + \|T\|_{H.S.}]}{\sqrt{2\pi}} \right) + \frac{\sqrt{\pi}}{12} M_2(h)\|\Lambda^{-1}\|_{op} \|\Sigma^{-1/2}\|_{op} E[\|W' - W\|_2^3]. \]

**Remark 3.1.16.** (i) Items (i) and (ii) of Remark 3.1.8 apply in an obvious way also to Theorem 3.1.15.

(ii) By Proposition 3.1.13 Condition (iii) in Theorem 3.1.15 is quite natural to be satisfied with a “small” matrix \( T \), if the approximation by a multivariate complex normal distribution makes sense.

**Proof of Theorem 3.1.15.** The proof is by reduction to the real version, Theorem 3.1.7. To this end, we will show, that the pair \((W_{R}, W'_{R})\) satisfies the assumptions of Theorem 3.1.7 with \( \Lambda_{R} \) for \( \Lambda \), \( \Sigma' := \frac{1}{2} \Sigma_{R} \) for \( \Sigma \), \( R_{R} \) for \( R \) and \( \frac{1}{2}(S_{R} + T_{R} J) \) for \( S \). Clearly, the pair \((W_{R}, W'_{R})\) is exchangeable and by rule (2) from Lemma 3.1.11 \( \Lambda_{R} \) is invertible. By rule (6) from Lemma 3.1.11 \( \Sigma' \) is positive semidefinite. As to Condition (3.29) we have by (i) that

\[ E\left[ W'_{R} - W_{R} \mid \mathcal{F} \right] = E\left[ W' - W \mid \mathcal{F} \right]_{R} = (-\Lambda W + R)_{R} = -\Lambda_{R} W_{R} + R_{R}. \]
3. Multivariate normal approximation and the classical compact groups

To prove (3.31), using representation (3.33), (ii) and (iii) we compute

\[ E\left[(W'_R - W_R)(W'_R - W_R)^T | \mathcal{F}\right] = E\left[(W' - W)(W' - W)^T | \mathcal{F}\right] \]

\[ = \frac{1}{2} E\left[(W' - W)(W' - W)^* | \mathcal{F}\right] + \frac{1}{2} E\left[(W' - W)(W' - W)^T | \mathcal{F}\right] \]

\[ = \frac{1}{2} \left(2\Lambda \Sigma + \tilde{S}\right) + \frac{1}{2} T_R J \]

Thus, we may apply Theorem 3.1.7 but with \(2d\) for \(d\). Recall that

\[ \tilde{Z} := \sqrt{2} Z_R \sim N_{2d}(0, I_{2d}) \]

and hence,

\[ (\Sigma^{'})^{1/2} \tilde{Z} = (\Sigma_R^{1/2}) Z_R = (\Sigma^{1/2} Z)_R = (Z_R) \]

Finally, we observe \(\|R_R\|_2 = \|R\|_2, \|W'_R - W_R\|_2 = \|W' - W\|_2\) and that by rules (7) and (8) from Lemma 3.1.11 we have \(\|(\Lambda_R)^{-1}\|_{\text{op}} = \|\Lambda^{-1}\|_{\text{op}}, \|S_R\|_{\text{h.s.}} = \sqrt{2}\|S\|_{\text{h.s.}}, \|T_R J\|_{\text{h.s.}} = \|T_R\|_{\text{h.s.}} = \sqrt{2}\|T\|_{\text{h.s.}}\) and \(\|(\Sigma')^{-1/2}\|_{\text{op}} = \sqrt{2}\|(\Sigma_R)^{-1/2}\|_{\text{op}} = \sqrt{2}\|\Sigma^{-1/2}\|_{\text{op}}\). Plugging this data into the bounds from Theorem 3.1.7 and using that

\[ \|S_R + T_R J\|_{\text{h.s.}} \leq \|S_R\|_{\text{h.s.}} + \|T_R J\|_{\text{h.s.}} = \sqrt{2}(\|S\|_{\text{h.s.}} + \|T\|_{\text{h.s.}}) \]

yields the assertion. \(\square\)

**Theorem 3.1.17.** Let \(W, W_t (t > 0)\) be \(C^d\)-valued \(L^2(P)\) random vectors on the same probability space \((\Omega, \mathcal{A}, P)\) such that for any \(t > 0\) the pair \((W, W_t)\) is exchangeable. Suppose there exist an invertible non-random matrix \(\Lambda \in C^{d \times d}\), a positive semidefinite matrix \(\Sigma \in C^{d \times d}\), a random vector \(R = (R_1, \ldots, R_d)^T\), random \(d \times d\)-matrices \(S\) and \(T\), a sub-\(\sigma\)-algebra \(\mathcal{F}\) of \(A\) with \(\sigma(W) \subseteq \mathcal{F}\) and a deterministic function \(s : (0, \infty) \to (0, \infty)\) such that the following three conditions are satisfied:

\[ \frac{1}{s(t)} E[|W_t - W|^{2}] \xrightarrow{t \to 0} -\Lambda W + R \text{ in } L^1(P). \]

(i)

\[ \frac{1}{s(t)} E[(W_t - W)(W_t - W)^* | \mathcal{F}] \xrightarrow{t \to 0} 2\Lambda \Sigma + S \text{ in } L^1(\|\cdot\|_{\text{h.s.}}). \]

(ii)

\[ \frac{1}{s(t)} E[(W_t - W)(W_t - W)^T | \mathcal{F}] \xrightarrow{t \to 0} T \text{ in } L^1(\|\cdot\|_{\text{h.s.}}). \]

(iii)

\[ \lim_{t \to 0} \frac{1}{s(t)} E\left[\|W_t - W\|^2 1_{\{\|W_t - W\|^2 > \varepsilon\}}\right] = 0 \text{ for each } \varepsilon > 0. \]

(iv)
3.1. Stein’s method for multivariate normal distributions

Then:

(a) For each $h \in C^2(\mathbb{R}^{2d})$ such that $E[|h(W)|] < \infty$ and $E[|h(Z_\Sigma)|] < \infty$,

\[
|E[h(W)] - E[h(Z_\Sigma)]| \leq \|\Lambda^{-1}\|_{op} \left( M_1(h)E[\|R\|_2] + \frac{1}{4\sqrt{2}} \tilde{M}_2(h)E[\|S\|_{H.S.} + \|T\|_{H.S.}] \right)
\]

\[
\leq \|\Lambda^{-1}\|_{op} \left( M_1(h)E[\|R\|_2] + \sqrt{\frac{d}{4}} \tilde{M}_2(h)E[\|S\|_{H.S.} + \|T\|_{H.S.}] \right).
\]

(b) If $\Sigma$ is actually positive definite, then

\[
d_W(W, Z_\Sigma) \leq \|\Lambda^{-1}\|_{op} \left( E[\|R\|_2] + \frac{1}{\sqrt{2\pi}} \|\Sigma^{-1/2}\|_{op} E[\|S\|_{H.S.} + \|T\|_{H.S.}] \right).
\]

Remark 3.1.18. Of course, Remark 3.1.10 also applies to Theorem 3.1.17.

Proof of Theorem 3.1.17. Again, the proof is by reduction to the real version, Theorem 3.1.9. This is very similar to the calculations in the proof of Theorem 3.1.15 but with $2d$ for $d$. Here, we show that the pairs $(W^*_R, (W_t)_R)$, $t > 0$, satisfy the assumptions of Theorem 3.1.9 with $\Lambda_R$ for $\Lambda$, $\Sigma' := \frac{1}{2}\Sigma_R$ for $\Sigma$, $R_R$ for $R$ and $\frac{1}{2}(S_R + T_R J)$ for $S$. Of course, the pair $(W^*_R, (W_t)_R)$ is exchangeable for each $t > 0$, $\Lambda_R$ is invertible by Lemma 3.1.11 (2) and by rule (6) from Lemma 3.1.11 $\Sigma'$ is positive semidefinite. By Condition (i) we have

\[
E[(W_t)_R - W_R | F] = E[W_t - W | F] \xrightarrow{t \to 0} (-\Lambda W + R)_R = -\Lambda_RW_R + R_R.
\]

Furthermore, using representation (3.33) and Assumptions (ii) and (iii) we compute

\[
E\left[ ((W_t)_R - W_R)(W_t)_R - W_R)^T | F \right] = E\left[ (W_t - W)_R (W_t - W)^T | F \right] = \frac{1}{2} E\left[ (W_t - W)^* (W_t - W)^T | F \right] J
\]

\[
\xrightarrow{t \to 0} \frac{1}{2} \left( 2\Lambda\Sigma + S \right)_R + \frac{1}{2} T_R J
\]

\[
= 2\Lambda_R\Sigma' + \frac{1}{2} (S_R + T_R J) \quad \text{in } L^1(\|\cdot\|_{H.S.})
\]
Finally, noting that $\| (W_t - W) \|_2 = \| W_t - W \|_2$ we see by (iv) that all the assumptions of Theorem 3.1.9 are satisfied. The computation of the other relevant quantities for the bounds is the same as in the proof of Theorem 3.1.15 and is therefore omitted.

\[ \square \]

**Remark 3.1.19.** In our “plug-in theorems”, Theorems 3.1.7, 3.1.9, 3.1.15 and 3.1.17, all of the bounds were given for “smooth” test functions $h$, meaning that they are differentiable or at least Lipschitz continuous. It should be noted that there exist theorems giving bounds for non-smooth test functions, e.g. indicator functions of convex sets or of left-infinite rectangles, which yield the multivariate Kolmogorov distance. In fact, even the first adoption of Stein’s method for multivariate normal approximation by Götze [Göt91] was about bounds in the multivariate CLT for iid summands in the non-smooth metric induced by indicator functions of convex sets. Since the appearance of [Göt91], (nearly) every adaption of Stein’s method for multivariate normal approximation for non-smooth test functions uses the technique proposed by Götze, which consists of considering a smoothed version $h_t$ (for small $t > 0$) of the test function $h$ and then using a smoothing inequality due to Bhattacharya and Rao [BRR86]. Although Götze obtains the optimal rate of convergence of $n^{-1/2}$ in the iid case, in more complicated situations this technique has the drawback that it usually only yields rates of order at best $n^{-1/2} \log n$. This is due to the bound $C \log(t^{-1})$ on the second derivative of the solution to the Stein equation for the smoothed test function $h_t$, which was already given in [Göt91] and which could not be improved yet. If the dimension $d = 1$, then already Götze has observed that this bound could be replaced by the constant 1, yielding optimal bounds in various theorems on univariate normal approximation. Within the exchangeable pairs approach for multivariate normal approximation, non-smooth test functions have been considered by Reinert and Röllin [RR09] and Eichelsbacher and Martschink [EM10].
3.2. Spectral properties of large random matrices from the classical compact groups

In the present section we deal with Haar distributed random matrices from one of the classical connected, compact Lie groups: the unitary group $U(n)$, the special orthogonal group $SO(n)$ and the unitary symplectic group $USp(2n)$. Of course, the groups $SO(n)$ and $U(n)$ are well known. To define the group $USp(2n)$, we first define the complex symplectic group $Sp(2n; \mathbb{C})$ to consist of all matrices $A \in \mathbb{C}^{2n \times 2n}$ such that

$$A^T \Omega A = \Omega,$$  \hspace{1cm} (3.38)

where $\Omega = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$, i.e. $Sp(2n; \mathbb{C})$ is the matrix group of isometries with respect to the bilinear form given by $\Omega$. This already implies that $Sp(2n; \mathbb{C})$ is a group with respect to matrix multiplication. Furthermore, $Sp(2n; \mathbb{C})$ is a non-compact subgroup of $SL(2n; \mathbb{C})$, the special linear group. In fact, from (3.38) one can easily see that $\det(A) = \pm 1$ for $A \in Sp(2n; \mathbb{C})$. Now, partitioning $A$ as

$$A = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix}$$

with $A_1, A_2, A_3, A_4 \in \mathbb{C}^{n\times n}$ and using block multiplication, we have from (3.38), that $A \in Sp(2n; \mathbb{C})$ if and only if

$$0 = A_1^T A_3 - A_3^T A_1 \iff A_1^T A_3 = A_3^T A_1 \quad \text{and} \quad I_n = A_1^T A_4 - A_3^T A_2.$$ \hspace{1cm} (3.39), (3.40)

Now, by the Schur complement and equations (3.39) and (3.40), we have

$$\det(A) = \det(A_1) \cdot \det(A_4 - A_3 A_1^{-1} A_2) = \det(A_1^T) \cdot \det(A_4 - A_3 A_1^{-1} A_2) = \det(A_1^T) \cdot \det(A_4 - A_3 A_1^{-1} A_2) = \det(A_1^T A_4 - A_3^T A_1 A_1^{-1} A_2) = \det(A_1^T A_4 - A_3^T A_2) = \det(I_n) = 1.$$

Thus, $Sp(2n; \mathbb{C})$ is indeed a subgroup of $SL(2n; \mathbb{C})$.

To see that $Sp(2n; \mathbb{C})$ is not compact, it suffices to note that for each complex number $c \neq 0$ the matrix $\begin{pmatrix} cI_n & 0 \\ 0 & c^{-1}I_n \end{pmatrix}$ belongs to $Sp(2n; \mathbb{C})$. However, if one defines $USp(2n) := Sp(2n; \mathbb{C}) \cap U(2n)$, then this defines a compact subgroup of $Sp(2n; \mathbb{C})$. 


3. Multivariate normal approximation and the classical compact groups

on which Haar measure may be normalized to a probability measure.

One aspect of random matrix theory concerns the spectral properties of Haar distributed random elements $M_n$ from one of the classical compact groups as $n \to \infty$. For example, a famous result due to Diaconis and Shahshahani [DS94] states that if $M_n$ is chosen from Haar measure on $U(n)$, $O(n)$ or $USp(2n)$, then for a fixed positive integer $d$, the distribution of the vector

$$(\text{Tr}(M_n), \text{Tr}(M_n^2), \ldots, \text{Tr}(M_n^d))^T$$

of traces of powers converges weakly to a $d$-dimensional (real or complex) normal distribution with diagonal covariance matrix. The original proof of this result used the method of moments in conjunction with exact formulae for the joint moments of the traces of powers, also due to Diaconis and Shahshahani [DS94]. Their proof of these moment formulae used representation theory and facts about special functions. Subsequently, several other proofs of these moment formulae were found (see, e.g., [Sto05], [HR03] and [PV04]. The last two references also cover the group $SO(n)$). Afterwards, the speed of convergence of $\text{Tr}(M_n^d)$ for a fixed $d$ to the univariate normal distribution was studied by several authors: In [Ste95], Stein used an infinitesimal version of his method of exchangeable pairs to prove that for the orthogonal group, the error with respect to the Kolmogorov distance decreases faster than any power of $n$. Later, in [Joh97], using Szegö's limit theorem for Toeplitz determinants, Johansson proved a conjecture by Diaconis, that in fact the convergence is exponentially fast for the groups $O(n)$ and $USp(2n)$, and is even super-exponentially fast for the unitary group $U(n)$. He also obtained these convergence rates for finite linear combinations of traces of various powers. Recently, in [Full01], Fulman proposed an approach to the speed of convergence of $\text{Tr}(M_n^d)$, based on combining Stein's method of exchangeable pairs with heat kernel techniques. Although his results on the speed of convergence are weaker than those of Stein and of Johansson, his theorems also apply to the case that the power $d$ grows with $n$ in such a way that still $d = o(n)$. Furthermore, his techniques seem more likely to be applicable in more general situations, as for example representations different from the standard representation or in symmetric space contexts.

The plan of this section is the following: In Subsection 3.2.1 we prove a quantitative version of the Diaconis-Shahshahani theorem by Stein's method of exchangeable pairs for multivariate normal approximation, see Subsection 3.1.2. This joint work with Michael Stolz has already been published as [DS11]. As a benefit from our concrete error bounds, we may allow the maximal power $d$ in the vector to depend on $n$. This last property enables us in Subsection 3.2.2 to prove a rate of convergence for the known Gaussian fluctuations of linear eigenvalue statistics.
3.2. Spectral properties of large random matrices from the classical compact groups

of large Haar distributed unitary matrices, if the test function considered is sufficiently smooth. This application is also joint work with Michael Stolz (see [DS12]).

3.2.1. Traces of powers

In this subsection we derive a rate of convergence in Wasserstein distance for the Diaconis-Shahshahani theorem. As was already indicated, we here consider the connected ones among the classical compact groups, namely the groups $U(n)$, $SO(n)$ and $USp(2n)$. Our approach uses the infinitesimal version of the exchangeable pairs technique for multivariate (real or complex) normal approximation as explained in Subsection 3.1.2 with the exchangeable pairs construction proposed by Fulman [Ful10]. This yields, to the best of our knowledge, the first rates of convergence result in the multivariate CLT for traces of powers on the classical groups, or, in any case, the first one that allows the maximal power $d$ to grow with $n$. Since Fulman’s approach consists in constructing the exchangeable pairs $(W, W_t)$ from Brownian motion on the group considered, this amounts to studying the heat kernel on this group and, in particular the action of the Laplacian on power sum symmetric polynomials. Crucial formulae for this action were given by Rains [Rai97a].

Since the proof of our main result is rather long and needs some auxiliary lemmas, we will first introduce the setting and state the theorem.

In the following, we let $M = M_n$ be a Haar distributed element of $K_n \in \{U(n), SO(n), USp(2n)\}$. For $d, r \in \mathbb{N}$ with $1 \leq r \leq d$ consider the $r$-dimensional (real or complex) random vector

$$W := (f_{d-r+1}(M), f_{d-r+2}(M), \ldots, f_d(M))^T,$$

where the $f_j, j = d - r + 1, \ldots, d$ are functions on $K_n$ with $f_j(M) := \text{Tr}(M^j)$ in the unitary case,

$$f_j(M) := \begin{cases} 
\text{Tr}(M^j), & j \text{ odd}, \\
\text{Tr}(M^j) - 1, & j \text{ even}
\end{cases}$$

in the orthogonal case and

$$f_j(M) := \begin{cases} 
\text{Tr}(M^j), & j \text{ odd}, \\
\text{Tr}(M^j) + 1, & j \text{ even}
\end{cases}$$

in the symplectic case. In the orthogonal and symplectic cases, let $Z := (Z_{d-r+1}, \ldots, Z_d)^T$ denote an $r$-dimensional real standard normal random vector. In the unitary case, $Z$ is defined as an $r$-dimensional standard complex normal
3. Multivariate normal approximation and the classical compact groups

vector, as introduced in Section 3.1. In all cases, we take $\Sigma$ to denote the diagonal matrix \( \text{diag}(d-r+1, d-r+2, \ldots, d) \) and write $Z_{\Sigma} := \Sigma^{1/2} Z$. Recall the Wasserstein distance, which was defined by equation (3.30).

**Theorem 3.2.1.** If $n \geq 2d$ in the unitary and symplectic cases, and $n \geq 4d + 1$ in the orthogonal case, the Wasserstein distance between $W$ and $Z_{\Sigma}$ is

$$d_{W}(W, Z_{\Sigma}) = O \left( \frac{\max \left\{ \frac{r^{7/2}}{(d-r+1)^{3/2}}, (d-r)^{3/2} \sqrt{r} \right\}}{n} \right).$$

In particular, for $r = d$ we have

$$d_{W}(W, Z_{\Sigma}) = O \left( \frac{d^{7/2}}{n} \right),$$

and for $r \equiv 1$

$$d_{W}(W, Z_{\Sigma}) = O \left( \frac{d^{3/2}}{n} \right).$$

If $1 \leq r = \lfloor cd \rfloor$ for $0 < c < 1$, then

$$d_{W}(W, Z_{\Sigma}) = O \left( \frac{d^{2}}{n} \right).$$

**Remark 3.2.2.** (i) The notation $O(\cdots)$ in the statement of Theorem 3.2.1 might be a bit ambiguous. For reasons of clearness we specify its exact meaning: For each of the sequences $(U(n))_{n \in \mathbb{N}}$, $(SO(n))_{n \in \mathbb{N}}$ and $(USp(2n))_{n \in \mathbb{N}}$ there exists a finite constant $C > 0$ such that for all positive integers $r$, $d$ and $n$ satisfying $1 \leq r \leq d$ and $n \geq 2d$ in the unitary and symplectic cases and $n \geq 4d + 1$ in the orthogonal case, respectively, we have

$$d_{W}(W, Z_{\Sigma}) \leq C \frac{\max \left\{ \frac{r^{7/2}}{(d-r+1)^{3/2}}, (d-r)^{3/2} \sqrt{r} \right\}}{n}.$$

(ii) The case $r \equiv 1$ means that one considers a (possibly shifted) single power $f_{d}(M)$. In his study [Ful10] of the univariate case, Fulman considers the random variable $\frac{f_{d}(M)}{\sqrt{d}}$ instead. By the scaling properties of the Wasserstein metric, the present result implies

$$d_{W} \left( \frac{f_{d}(M)}{\sqrt{d}}, N(0, 1) \right) = O \left( \frac{d}{n} \right)$$

128
in the special orthogonal and in the unitary symplectic cases and the same result but with $N(0,1)$ replaced by $N_{1,C}(0,1)$ in the unitary case. So in the one-dimensional special case we recover the rate of convergence that was obtained by Fulman [Ful10], albeit in Wasserstein rather than Kolmogorov distance, which he considered.

(iii) One might wonder if the CLT for $d^{-1/2} \operatorname{Tr}(M^d)$ which by (ii) holds for $d = o(n)$ is still valid if this condition is abandoned. The answer to this question is negative, since it was proved by Rains [Rai97b] that, in the unitary case, for $d \geq n$ the eigenvalues of $M^d$ are independent and uniformly distributed on $S^1$ and, hence, by the CLT in $C \cong \mathbb{R}^2$ the statistic $\operatorname{Tr}(M^d)$ approaches a complex normal random variable after dividing it by $\sqrt{n}$ instead of $\sqrt{d}$. Since Rains result extends (with some caveats) to general compact Lie groups, a similar remark applies to the groups $SO(n)$ and $USp(2n)$.

Now, let us turn to the proof of Theorem 3.2.1. First, we construct a family $(W,W_t)$, $t > 0$, of exchangeable pairs along the lines of Fulman’s univariate approach in [Ful10].

**Construction 3.2.3.** Let $(M_t)_{t \geq 0}$ be Brownian motion on $K_n$ and take $M := M_0$ to be Haar distributed. Brownian motion being reversible with respect to Haar measure, $(M,M_t)$ is an exchangeable pair for each $t > 0$. Now, let the functions $f_j$, $j = d - r + 1, \ldots, r$, and the vector $W = (f_{d-r+1}(M), \ldots, f_d(M))^T$ be given as above. For $t > 0$ we define the vector $W_t := (f_{d-r+1}(M_t), \ldots, f_d(M_t))^T$. Then, clearly, also $(W,W_t)$ is exchangeable for each $t > 0$. To be more specific, denote by $\Delta = \Delta_{K_n}$ the Laplace-Beltrami operator (the Laplacian) on $K_n$. Then, $(M_t)_{t \geq 0}$ is the diffusion on $K_n$ with infinitesimal generator $\Delta$ and Haar measure as initial distribution. Reversibility then follows from general theory (see [Hel00, Section II.2.4], [IW89, Section V.4], for the relevant facts). Let $(T_t)_{t \geq 0}$, often symbolically written as $(e^{t\Delta})_{t \geq 0}$, be the corresponding semigroup of Markov transition operators on $C^2(K_n)$. Thus, for $f \in C^2(K_n)$ we have

$$E[f(M_t) \mid M] = (T_t f)(M) \text{ P-a.s.}$$

**Lemma 3.2.4 (Regression lemma).** Let $f \in C^2(K_n)$. Then,

$$E[f(M_t) \mid M] = f(M) + t(\Delta f)(M) + O(t^2).$$

**Proof.** Note that the mapping $[0, \infty) \times K_n \ni (t,g) \mapsto (T_t f)(g)$ satisfies the heat equation

$$\frac{\partial}{\partial t} (T_t f)(g) = (\Delta T_t f)(g).$$
3. Multivariate normal approximation and the classical compact groups

So, by (3.41) and a Taylor expansion in $t$ at 0 we obtain

$$ E[f(M_t) \mid M] = (T_t f)(M) = (T_0 f)(M) + t \left( \frac{\partial}{\partial t} (T_t f)(M) \right)_{t=0} + O(t^2) $$

$$ = f(M) + t(\Delta f)(M) + O(t^2). $$

Remark 3.2.5. The $O(t^2)$ term in Lemma 3.2.4 deserves some explanation. By continuity of derivatives and the compactness of $K_n$, there is an absolute constant $C \in (0, \infty)$ such that the term $O(t^2)$ is dominated by $Ct^2$ no matter what the (random) value of $M$ is. Note that this also implies, that

$$ \frac{1}{t} E[f(M_t) - f(M) \mid M] \xrightarrow{t \to 0} (\Delta f)(M) \text{ a.s. and in } L^1(P). $$

Arguments of this type will occur frequently in what follows, usually without further mention.

Having discussed the family of exchangeable pairs, we now come to power sum symmetric polynomials and the action of the Laplacian on them. Let $X_1, \ldots, X_n$ be algebraic indeterminates and consider a finite family $\lambda = (\lambda_1, \ldots, \lambda_r)$ of positive integers (which could be ordered to obtain an integer partition). Then, the power sum symmetric polynomial indexed by $\lambda$ is defined by

$$ p_\lambda = \prod_{k=1}^r \sum_{j=1}^n X_j^{\lambda_k} = \prod_{k=1}^r p_{\lambda_k} \in \mathbb{C}[X_1, \ldots, X_n], $$

where $\mathbb{C}[X_1, \ldots, X_n]$ denotes the ring of polynomials in $n$ indeterminates over the field $\mathbb{C}$. Since $\mathbb{C}$ is an infinite field, we may identify polynomials with coefficients in $\mathbb{C}$ with polynomial functions and, hence, each $p \in \mathbb{C}[X_1, \ldots, X_n]$ gives rise to the function $\mathbb{C}^n \ni (z_1, \ldots, z_n) \mapsto p(z_1, \ldots, z_n) \in \mathbb{C}$. If $A \in \mathbb{C}^{n \times n}$ with not necessarily distinct eigenvalues $c_1, \ldots, c_n$, then we write $p_\lambda(A)$ in the place of $p_\lambda(c_1, \ldots, c_n)$. This is well-defined by the symmetry of $p_\lambda$ in $X_1, \ldots, X_n$. Then, we have the identity

$$ p_\lambda(A) = \prod_{k=1}^r \text{Tr}(A^{\lambda_k}) = \prod_{k=1}^r p_{\lambda_k}(A). $$

If $A \in \mathbb{C}^{n \times n}$ is invertible, then $p_\lambda(A)$ may also be defined for $\lambda = (\lambda_1, \ldots, \lambda_r) \in \mathbb{Z}^r$ by letting

$$ p_0(A) := \text{Tr}(I_n) = n, \quad p_{-k}(A) := p_k(A^{-1}) \text{ for } k \geq 1 \text{ and } p_\lambda(A) := \prod_{k=1}^r p_{\lambda_k}(A). $$
3.2. Spectral properties of large random matrices from the classical compact groups

Hence, for each \( \lambda \in \mathbb{Z}^r \), we may view \( p_\lambda \) as a function on \( K_n \). This function is complex-valued in the case \( K_n = U(n) \) and is real-valued if \( K_n \) is \( SO(n) \) or \( USp(2n) \). In the latter case, this follows, since the eigenvalues of \( A \in USp(2n) \) come in conjugate pairs. By our definitions one furthermore has the identities

\[
p_{-\lambda} = \overline{p_\lambda} \text{ in the unitary case},
p_{-\lambda} = p_\lambda \text{ in the special orthogonal case and}
p_{-\lambda} = p_\lambda \text{ in the unitary symplectic case}.
\]

For the unitary group, this follows from the fact that if \( U \in U(n) \), then \( U^{-1} = U^* \) and hence, for \( k \in \mathbb{Z} \),

\[
p_{-k}(U) = p_k(U^{-1}) = \text{Tr}((U^{-1})^k) = \text{Tr}((U^*)^k) = \text{Tr}((U^k)^*) = \overline{\text{Tr}(U^k)} = \overline{p_k}(U).
\]

For the special orthogonal group, the formula follows similarly from \( A^{-1} = A^T \) if \( A \in SO(n) \) and for the unitary symplectic group this follows from \( U^{-1} = U^* \) if \( U \in USp(2n) \) and from the fact, that the trace is real-valued on this group. The action of the Laplacian \( \Delta_{K_n} \) on these functions \( p_\lambda \) was studied by Rains [Rai97a] and Lévy [Lév08]. The following lemmas contain their formulae in the special cases that we need. For ease of notation, in the following we will constantly write \( p_{j,k} \) in the place of \( p_{(j,k)} \) and so on.

**Lemma 3.2.6.** For the Laplacian \( \Delta = \Delta_{U(n)} \) on \( U(n) \) and any nonnegative integers \( j, k \) one has

\[
\Delta p_j = -np_j - j \sum_{l=1}^{j-1} p_{j-l}.
\]

\[
\Delta p_{j,k} = -n(j+k)p_{j,k} - 2jkp_{j+k} - jpk \sum_{l=1}^{j-1} p_{j-l} - kp_j \sum_{l=1}^{k-1} p_{k-l}.
\]

\[
\Delta (p_j p_k) = 2jkp_{j-k} - n(j+k)p_{j+k} - jpk \sum_{l=1}^{j-1} p_{j-l} - kp_j \sum_{l=1}^{k-1} p_{k-l}.
\]

**Lemma 3.2.7.** For the Laplacian \( \Delta = \Delta_{SO(n)} \) on \( SO(n) \) and any nonnegative
3. Multivariate normal approximation and the classical compact groups

integers \( j, k \),

\[
\Delta p_j = -\frac{(n-1)}{2} j p_j - \frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l} + \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}. \tag{i}
\]

\[
\Delta p_{j,k} = -\frac{(n-1)(j+k)}{2} p_{j,k} - \frac{j}{2} p_k \sum_{l=1}^{j-1} p_{l,j-l} - \frac{k}{2} p_j \sum_{l=1}^{k-1} p_{l,k-l} - jk p_{j+k} + \frac{j}{2} p_k \sum_{l=1}^{j-1} p_{l,j-l} + \frac{k}{2} p_j \sum_{l=1}^{k-1} p_{l,k-l} + jk p_{j-k}. \tag{ii}
\]

**Lemma 3.2.8.** For the Laplacian \( \Delta = \Delta_{USp(2n)} \) on \( USp(2n) \) and any nonnegative integers \( j, k \) we have,

\[
\Delta p_j = -\frac{(2n+1)}{2} j p_j - \frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l} - \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}. \tag{i}
\]

\[
\Delta p_{j,k} = -\frac{(2n+1)}{2} (j+k) p_{j,k} - jk p_{j+k} - \frac{j}{2} p_k \sum_{l=1}^{j-1} p_{l,j-l} - \frac{k}{2} p_j \sum_{l=1}^{k-1} p_{l,k-l} - jk p_{j+k} + \frac{j}{2} p_k \sum_{l=1}^{j-1} p_{l,j-l} + \frac{k}{2} p_j \sum_{l=1}^{k-1} p_{l,k-l} + jk p_{j-k}. \tag{ii}
\]

In our proof of Theorem 3.2.1 we will need to integrate certain \( p_{\lambda} \) over the group \( K_n \). Explicit formulae for these integrals, which we now recall, are due to Diaconis and Shahshahani [DS94].

Let \( a = (a_1, \ldots, a_r) \) and \( b = (b_1, \ldots, b_s) \) be tuples of nonnegative integers. We define functions \( f_a : \mathbb{N}_{\geq 1} \to \mathbb{R} \) by

\[
f_a(j) := \begin{cases} 
1, & \text{if } a_j = 0 \\
0, & \text{if } ja_j \text{ is odd} \\
ja_j/2(a_j - 1)!! & , & \text{if } j \text{ is odd and } a_j \geq 2 \text{ is even} \\
1 + \sum_{k=1}^{\lfloor a_j/2 \rfloor} j^k (a_j/2k) (2k-1)!! & , & \text{if } j \text{ is even and } a_j \geq 1 .
\end{cases}
\]

Here we have used the shorthand notation \( (2k-1)!! = (2k-1)(2k-3) \ldots 3 \cdot 1 \). Furthermore, we will write

\[
k_a := \sum_{j=1}^{r} ja_j , \ k_b := \sum_{j=1}^{s} jb_j , \text{ and } \eta_j := \frac{1 + (-1)^j}{2} = \begin{cases} 
1, & \text{if } j \text{ is even} \\
0, & \text{if } j \text{ is odd}.
\end{cases}
\]
3.2. Spectral properties of large random matrices from the classical compact groups

**Lemma 3.2.9.** Let $M = M_n$ be a Haar-distributed element of $U(n)$, $Z_1, \ldots, Z_r$ iid complex standard normals. Then, if $k_a \neq k_b$,

$$E \left[ \prod_{j=1}^{r} (\text{Tr}(M^j))^{a_j} \prod_{j=1}^{s} (\text{Tr}(M^j))^{b_j} \right] = 0.$$  

If $k_a = k_b$ and $n \geq k_a$, then the following formula holds:

$$E \left[ \prod_{j=1}^{r} (\text{Tr}(M^j))^{a_j} \prod_{j=1}^{s} (\text{Tr}(M^j))^{b_j} \right] = E \left[ \prod_{j=1}^{r} (\sqrt{j}Z_j)^{a_j} \prod_{j=1}^{s} (\sqrt{j}Z_j)^{b_j} \right] = \delta_{a,b} \prod_{j=1}^{r} j^{a_j} a_j!$$

**Lemma 3.2.10.** If $M = M_n$ is a Haar-distributed element of $SO(n)$, $n \geq k_a + 1$ and $Z_1, \ldots, Z_r$ are iid real standard normals, then

$$E \left[ \prod_{j=1}^{r} (\text{Tr}(M^j))^{a_j} \right] = E \left[ \prod_{j=1}^{r} (\sqrt{j}Z_j + \eta_j)^{a_j} \right] = \prod_{j=1}^{r} f_a(j).$$

**Lemma 3.2.11.** If $M = M_{2n}$ is a Haar-distributed element of $USp(2n)$, $2n \geq k_a - 1$ and $Z_1, \ldots, Z_r$ are iid real standard normals, then

$$E \left[ \prod_{j=1}^{r} (\text{Tr}(M^j))^{a_j} \right] = E \left[ \prod_{j=1}^{r} (\sqrt{j}Z_j - \eta_j)^{a_j} \right] = \prod_{j=1}^{r} (-1)^{(j-1)a_j} f_a(j).$$

**Remark 3.2.12.** Proofs of Lemmas 3.2.9, 3.2.10 and 3.2.11 can be found in [DS94], [DE01], [HR03], [PV04] and [Sto05]. Note that the moment formulae from Lemmas 3.2.9, 3.2.10 and 3.2.11 just state, that the vector of traces of powers of a random element from one of the classical compact groups has exactly the same joint moments as a suitable normally distributed vector, if the group index $n$ is large enough.

The first claim of Lemma 3.2.9 can easily be proved by multiplying $M$ from the left by the non-random unitary matrix $e^{i\theta}I_n$, where $\theta \in \mathbb{R}$ is not an integer multiple of $2\pi$ and using the invariance of Haar measure.

The following simple lemma will be useful for bounding partial sums of integer powers.

**Lemma 3.2.13.** Let $a, r, p$ be nonnegative integers, $r > 0$. Then

$$\sum_{k=a}^{a+r-1} k^p \leq \frac{2p+1}{p+1} \max\{ra^p, r^{p+1}\}.$$
3. Multivariate normal approximation and the classical compact groups

Proof. Since the function \( f : [0, \infty) \rightarrow \mathbb{R} \) with \( f(x) = x^p \) is nondecreasing,

\[
\sum_{k=a}^{a+r-1} k^p \leq \int_{a}^{a+r} x^p dx = \frac{(a + r)^{p+1} - a^{p+1}}{p + 1}
\]

\[
= \frac{r}{p + 1} \sum_{k=0}^{p} \frac{(p + 1)}{k} a^k r^{p-k} \leq \frac{r \max\{a^p, r^p\}}{p + 1} \sum_{k=0}^{p} \frac{(p + 1)}{k}
\]

\[
\leq \frac{2^{p+1}}{p + 1} \max\{r a^p, r^{p+1}\}.
\]

Now we are in the position to prove Theorem 3.2.1. This will be done for each of the groups \( U(n) \), \( SO(n) \) and \( USp(2n) \) separately by checking the conditions of either Theorem 3.1.17 in the unitary case or of Theorem 3.1.9 in the orthogonal and symplectic cases and using the respective bounds. In every case we take \( \mathcal{F} := \sigma(M) \) and \( s(t) := t \).

Proof of Theorem 3.2.1 for \( K_n = U(n) \). In this case we have to check the conditions (i)-(iv) of Theorem 3.1.17 for the family of exchangeable pairs \( (W, W_t), t > 0 \), from Construction 3.2.3. In order to verify Condition (i), we will need the following lemma.

**Lemma 3.2.14.** For all \( j = d - r + 1, \ldots, d \),

\[
E[W_{i,j} - W_j | M] = E[p_j(M_i) - p_j(M) | M] = t\left(-nj W_j + R_j + O(t)\right),
\]

where \( R_j := -j \sum_{l=1}^{j-1} p_{i,j-l}(M) \).

Proof. From Lemma 3.2.4 and Lemma 3.2.6 (i) we have

\[
E[p_j(M_i) | M] = p_j(M) + t(\Delta p_j)(M) + O(t^2)
\]

\[
= p_j(M) - tnjp_j(M) - tj \sum_{l=1}^{j-1} p_{i,j-l}(M) + O(t^2),
\]

implying the result.

\[\square\]
3.2. Spectral properties of large random matrices from the classical compact groups

From Lemma 3.2.14 and Remark 3.2.5 we conclude, that as $t \to 0$,
\[
\frac{1}{t} E[W_t - W | M] \to -\Lambda W + R \text{ a.s. and in } L^1(P),
\]
where $\Lambda := \text{diag}(n_j : j = d - r + 1, \ldots, d)$ and $R := (R_{d-r+1}, \ldots, R_d)^T$.

To verify Conditions (ii) and (iii) of Theorem 3.1.17, we first prove the following lemma.

**Lemma 3.2.15.** For all nonnegative integers $j, k$, one has
\[
E[(p_j(M_t) - p_j(M))(p_k(M_t) - p_k(M)) | M] = -2tjk p_{j+k}(M) + O(t^2) \quad (i)
\]
\[
E[(p_j(M_t) - p_j(M))(\overline{p_k(M)} - p_k(M)) | M] = 2tjk p_{j-k}(M) + O(t^2) \quad (ii)
\]

**Proof.** By well known properties of conditional expectation,
\[
E[(p_j(M_t) - p_j(M))(p_k(M_t) - p_k(M)) | M] = E[p_j(M_t)p_k(M_t) | M] - p_j(M)E[p_k(M_t) | M] - p_k(M)E[p_j(M_t) | M] + p_j(M)p_k(M).
\]

Applying Lemmas 3.2.4 and 3.2.6 (ii) to the first term yields
\[
E[p_{j,k}(M_t) | M] = p_{j,k}(M) + t(\Delta p_{j,k}(M)) + O(t^2)
\]
\[
= p_{j,k}(M) + t\left(-n(j + k)p_{j,k}(M) - 2jkp_{j+k}(M) - jp_k(M)\sum_{l=1}^{j-1} p_{l,j-l}(M) \right.
\]
\[
- k p_j(M)\sum_{l=1}^{k-1} p_{l,k-l}(M) \bigg) + O(t^2).
\]

Analogously, for the second term,
\[
p_j(M)E[p_k(M_t) | M] = p_j(M)p_k(M) + t(\Delta p_k)(M) + O(t^2)
\]
\[
= p_{j,k}(M) + tp_j(M)\left(-np_k(M) - k\sum_{l=1}^{k-1} p_{l,k-l}(M) \bigg) + O(t^2)
\]
\[
= p_{j,k}(M) - tnp_{j,k}(M) - kp_j(M)\sum_{l=1}^{k-1} p_{l,k-l}(M) + O(t^2),
\]
and by symmetry
3. Multivariate normal approximation and the classical compact groups

\[ p_k(M)E[p_j(M_i)\mid M] = p_{j,k}(M) - tn_j p_{j,k}(M) - t j p_k(M) \sum_{l=1}^{j-1} p_{l,j-l}(M) + O(t^2). \]

Summing up, one obtains

\[ E[(p_j(M_i) - p_j(M))(p_k(M_i) - p_k(M))\mid M] = -2t j k p_{j+k}(M) + O(t^2), \]

proving assertion (i). For the second assertion, we compute analogously

\[
E[(p_j(M_i) - p_j(M))(\overline{p_k}(M_i) - \overline{p_k}(M))\mid M] = E[p_j(\overline{p_k}(M_i))\mid M] - p_j(M)E[\overline{p_k}(M_i)\mid M] - \overline{p_k}(M)E[p_j(M_i)\mid M] + p_j(M)\overline{p_k}(M),
\]

and we have by Lemma 3.2.6 (iii)

\[
E[p_j(\overline{p_k}(M_i))\mid M] = p_j(M)\overline{p_k}(M) + t \Delta(p_j\overline{p_k})(M) + O(t^2)
\]

\[
= p_j\overline{p_k}(M) + t \left( 2j k p_{j-k}(M) - n(j + k)p_j\overline{p_k}(M) 
- j \overline{p_k}(M) \sum_{l=1}^{j-1} p_{l,j-l}(M) - k p_j(M) \sum_{l=1}^{k-1} \overline{p_k}_{l,k-l}(M) \right) + O(t^2)
\]

as well as

\[
p_j(M)E[\overline{p_k}(M_i)\mid M] = p_j(M)\left( \overline{p_k}(M) + t(\Delta\overline{p_k})(M) + O(t^2) \right)
\]

\[
= p_j(M)\overline{p_k}(M) + t p_j(M) \left( -n k p_j(M) - k \sum_{l=1}^{k-1} \overline{p_k}_{l,k-l}(M) \right) + O(t^2)
\]

and

\[
\overline{p_k}(M)E[p_j(M_i)\mid M] = \overline{p_k}(M) \left( p_j(M) + t(\Delta p_j)(M) + O(t^2) \right)
\]

\[
= p_j\overline{p_k}(M) + t \overline{p_k}(M) \left( -n j p_j(M) - j \sum_{l=1}^{j-1} p_{l,j-l}(M) \right) + O(t^2).\]

Summing up, one has

\[ E[(p_j(M_i) - p_j(M))(\overline{p_k}(M_i) - \overline{p_k}(M))\mid M] = 2t j k p_{j-k} + O(t^2). \]

\[ \square \]
3.2. Spectral properties of large random matrices from the classical compact groups

Now we are in a position to identify the random matrices $S, T$ from Theorem 3.1.17. By Lemma 3.2.15 (i), \( \frac{1}{t} E[(W_t - W)(W_t - W)^T]\) converges almost surely, and in \( L^1(\|\cdot\|_{H,S})\), to \( T = (t_{jk})_{j,k=d-r+1,...,d} \), where \( t_{jk} = -2j kp_j + k(M) \) for \( j, k = d - r + 1, \ldots, d \). Observing that \( \Lambda \Sigma = \text{diag}(nu^2) : j = d - r + 1, \ldots, d \), one has from Lemma 3.2.15 (ii) that \( \frac{1}{t} E[(W_t - W)(W_t - W)^T]\) converges almost surely, and in \( L^1(\|\cdot\|_{H,S})\), to \( 2 \Lambda \Sigma + S \), where \( S = (s_{jk})_{j,k=d-r+1,...,d} \) is given by

\[
s_{jk} = \begin{cases} 0, & \text{if } j = k \\ 2 j kp_{j-k}(M), & \text{if } j \neq k. \end{cases}
\]

Next, we will verify Condition (iv), using Remark 3.1.10. Specifically, we will show that \( E[\|W_t - W\|^2] = O(t^{3/2}) \). Since, by Hölder’s inequality,

\[
E[\|W_t - W\|^2] = E\left[\left( \sum_{j=d-r+1}^d (W_{t,j} - W_j)^2 \right)^{3/2}\right]
\]

\[
= E\left[\left( \sum_{j,k,l=d-r+1}^d (W_{t,j} - W_j)^2 (W_{t,k} - W_k)^2 (W_{t,l} - W_l)^2 \right)^{1/2}\right]
\]

\[
\leq \sum_{j,k,l=d-r+1}^d E[|W_{t,j} - W_j|^3] E[|W_{t,k} - W_k|^3] E[|W_{t,l} - W_l|^3]^{1/3},
\]

it suffices to show that \( E[|W_{t,j} - W_j|^3] = O(t^{3/2}) \) for all \( j = d - r + 1, \ldots, d \). This in turn follows from the next lemma, since

\[
E[|W_{t,j} - W_j|^3] \leq \left( E[|W_{t,j} - W_j|^4] E[|W_{t,j} - W_j|^2] \right)^{1/2}.
\]

**Lemma 3.2.16.** For \( j = d - r + 1, \ldots, d, \ n \geq 2d, \)

\[
E[|W_{t,j} - W_j|^2] = 2j^2 nt + O(t^2), \quad \text{(i)}
\]

\[
E[|W_{t,j} - W_j|^4] = O(t^2). \quad \text{(ii)}
\]

**Proof.** By Lemma 3.2.15 (ii),

\[
E[|W_{t,j} - W_j|^2] = E[(p_j(M_t) - p_j(M)) (p_j(M_t) - p_j(M))] = E\left[ E[(p_j(M_t) - p_j(M)) (p_j(M_t) - p_j(M)) | M] \right]
\]

\[
= E[2t j^2 n + O(t^2)] = 2t j^2 n + O(t^2),
\]

137
establishing (i). Turning to (ii), one calculates that

\[
E[|W_{t,j} - W_j|^4] = E[(W_{t,j} - W_j)^2 \cdot (W_{t,j} - W_j)^2]
\]

\[
= E[(p_j(M_t) - p_j(M))^2 \cdot (\overline{\alpha_j}(M_t) - \overline{\alpha_j}(M))^2]
\]

\[
= E[p_j^2(M_t) - 2p_j(M)p_j(M_t) + p_j^2(M)](\overline{\alpha_j}^2(M_t) - 2\overline{\alpha_j}(M)\overline{\alpha_j}(M_t) + \overline{\alpha_j}^2(M))
\]

\[
= E[p_j^2(M_t)\overline{\alpha_j}^2(M_t)] - 2E[p_j^2(M_t)\overline{\alpha_j}(M_t)\overline{\alpha_j}(M)] - 2E[p_j^2(M_t)p_j(M)p_j(M_t)]
\]

\[
+ E[p_j^2(M_t)\overline{\alpha_j}^2(M)] + E[\overline{\alpha_j}^2(M_t)p_j^2(M)] + 4E[p_j(M_t)\overline{\alpha_j}(M_t)p_j(M)\overline{\alpha_j}(M)]
\]

\[
- 2E[p_j^2(M)\overline{\alpha_j}(M)\overline{\alpha_j}(M_t)] - 2E[\overline{\alpha_j}^2(M)p_j(M)p_j(M_t)] + E[p_j^2(M)p_j^2(\overline{\alpha_j})]
\]

\[= S_1 - 2S_2 - 2S_3 + S_4 + S_5 + 4S_6 - 2S_7 - 2S_8 + S_9.
\]

By exchangeability, we have \( S_1 = S_9, \ S_3 = S_7, \ S_4 = S_5, \ S_7 = S_2, \ S_8 = \overline{S_7} = S_3. \)

Now, for \( n \geq 2d, \) i.e., large enough for the moment formulae from Lemma 3.2.9 to apply for all \( j = d - r + 1, \ldots, d, \) we have by Lemmas 3.2.4 and 3.2.6 (i)

\[
S_1 = 2j^2,
\]

\[
S_8 = E[\overline{\alpha_j}^2(M)p_j(M)p_j(M_t)] = E[\overline{\alpha_j}^2(M)p_j(M)|M]
\]

\[
= E[\overline{\alpha_j}^2(M)p_j(M_t)(p_j(M) + t(\Delta p_j)(M) + O(t^2))]
\]

\[
= E[\overline{\alpha_j}^2(M)p_j^2(M)] + E[\overline{\alpha_j}^2(M)p_j(M)O(t^2)]
\]

\[
+ tE[\overline{\alpha_j}^2(M)p_j(M)\overline{\alpha_j}(M) - \overline{\alpha_j}(M)\overline{\alpha_j}(M) - j \sum_{l=1}^{j-1} p_t,_{j-l}(M))]
\]

\[
= 2j^2 + O(t^2) - tn_j2j^2 - tj \sum_{l=1}^{j-1} E[\overline{\alpha_j}^2(M)p_j(M)p_t,_{j-l}(M)]
\]

\[
= 2j^2 - 2tn_j^3 + O(t^2).
\]

Hence, \( S_3 = S_8 = \overline{S_7} = \overline{S_5} = 2j^2 - 2tn_j^3 + O(t^2). \) On the other hand, again by Lemmas 3.2.4 and 3.2.6 (ii)

138
3.2. Spectral properties of large random matrices from the classical compact groups

\[ S_4 = E[p_j^2(M_t)\overline{p}_j^2(M)] = E[\overline{p}_j^2(M)E[p_j^2(M_t)|M]] \]
\[ = E[\overline{p}_j^2(M)(p_j^2(M) + t(\Delta p_j^2)(M) + O(t^2))] = E[\overline{p}_j^2(M)p_j^2(M)] + O(t^2) \]
\[ + tE[\overline{p}_j^2(M)(-2npj^2(M) - 2j^2p_j(M) - 2j\sum_{l=1}^{j-1}p_{l,j-l}(M))] \]
\[ = 2j^2 - 4ntj^3 - 2t^2E[\overline{p}_j^2(M)p_{2j}(M)] \]
\[ - 2tj \sum_{l=1}^{j-1}E[p_j^2(M)p_{l,j-l}(M)] + O(t^2) \]
\[ = 2j^2 - 4ntj^3 + O(t^2). \]

Finally, by Lemmas 3.2.4 and 3.2.6 (iii) we have

\[ S_6 = E[p_j(M_t)p_j(M)\overline{p}_j(M)] = E[p_j(M)\overline{p}_j(M)E[p_j(M_t)\overline{p}_j(M_t) | M]] \]
\[ = E[p_j(M)\overline{p}_j(M)(p_j(M)\overline{p}_j(M) + t(\Delta p_j\overline{p}_j)(M) + O(t^2))] \]
\[ = E[p_j^2(M)\overline{p}_j^2(M)] + O(t^2) + tE[p_j(M)\overline{p}_j(M)] \]
\[ \left(2j^2n - 2ntj^3 - j\sum_{l=1}^{j-1}\overline{p}_{l,j-l}(M) - j\overline{p}_j(M)\sum_{l=1}^{j-1}p_{l,j-l}(M) \right) \]
\[ = 2j^2 + 2tj^3nE[p_j(M)\overline{p}_j(M)] - 2ntjE[p_j^2(M)\overline{p}_j^2(M)] \]
\[ - tj \sum_{l=1}^{j-1}E[p_j^2(M)p_{l,j-l}(M)\overline{p}_j(M)] \]
\[ - tj \sum_{l=1}^{j-1}E[p_j(M)p_{l,j-l}(M)\overline{p}_j^2(M)] + O(t^2) \]
\[ = 2j^2 + 2tnj^3 + 4ntj^3 + O(t^2) = 2j^2 + 2tnj^3 + O(t^2). \]

Putting the pieces together,

\[ E[|W_{i,j} - W_j|^4] = 2 \cdot 2j^2 - 8(2j^2 - 2tnj^3) + 2(2j^2 - 4tnj^3) \]
\[ + 4(2j^2 - 2tnj^3) + O(t^2) \]
\[ = j^2(4 - 16 + 4 + 8) + tnj^3(16 - 8 - 8) + O(t^2) = O(t^2), \]

as asserted.

\[ \square \]
3. Multivariate normal approximation and the classical compact groups

With the conditions of Theorem 3.1.17 in place, we have

\[
d_W(W, Z, \Sigma) \leq \|\Lambda^{-1}\|_{\text{op}} \left( E[\|R\|_2^2] + \frac{1}{\sqrt{2\pi}} \|\Sigma^{-1/2}\|_{\text{op}} E[\|S\|_{\text{H.S.}} + \|T\|_{\text{H.S.}}] \right). \tag{3.42}
\]

To bound the quantities on the right hand side of (3.42), we first observe that

\[
\|\Lambda^{-1}\|_{\text{op}} = \frac{1}{n(d-r+1)} \quad \text{and} \quad \|\Sigma^{-1/2}\|_{\text{op}} = \frac{1}{\sqrt{d-r+1}}.
\]

Now, for \( n \geq d \), Lemma 3.2.9 implies

\[
E[\|R\|_2^2] = \sum_{j=d-r+1}^d E[R_j R_j] = \sum_{j=d-r+1}^d j^2 \sum_{l,m=1}^{j-1} E[p_{l,j-l}(M)p_{m,j-m}(M)].
\]

For \( n \geq d \), Lemma 3.2.9 implies

\[
E[p_{l,j-l}(M)p_{m,j-m}(M)] = \delta_{l,m} E[p_{l,j-l}(M)p_{l,j-l}(M)]
\]

and

\[
E[p_{l,j-l}(M)p_{l,j-l}(M)] = \begin{cases} 2l(j-l), & l = \frac{j}{2} \\ l(j-l), & \text{otherwise} \end{cases} \leq 2l(j-l).
\]

Hence, using Lemma 3.2.13

\[
E[\|R\|_2^2] \leq 2 \sum_{j=d-r+1}^d j^2 \sum_{l=1}^{j-1} l(j-l) = 2 \sum_{j=d-r+1}^d j^2 j (j^2 - 1) / 6
\]

\[
\leq \frac{1}{3} \sum_{j=d-r+1}^d (j^5 - j^3) \leq \sum_{j=d-r+1}^d j^5
\]

\[
\leq \frac{26}{3 \cdot 6} \max\{(r-1)(d-r+1)^5, (r-1)^6\}
\]

\[
= O\left(\max(r^6, r(d-r)^5)\right),
\]

and thus,

\[
E[\|R\|_2] \leq \left( E[\|R\|_2^2] \right)^{1/2} = O\left(\max(r^3, \sqrt{r(d-r)^5/2})\right).
\]

Furthermore, again by Lemmas 3.2.9 and 3.2.13, for \( n \geq d \),
3.2. Spectral properties of large random matrices from the classical compact groups

\[ E[\|S\|_{\text{H.S.}}^2] = \sum_{j,k=d-r+1}^{d} E[|s_{jk}|^2] = 8 \sum_{d-r+1 \leq k < j \leq d} j^2 k^2 E[p_{j-k}(M)p_{j-k}(M)] \]

\[ = 8 \sum_{d-r+1 \leq k < j \leq d} j^2 k^2 (j - k) \]

\[ = 8 \sum_{d-r+1 \leq k < j \leq d} j^3 k^2 - 8 \sum_{d-r+1 \leq k < j \leq d} j^2 k^3 \]

\[ \leq 8 \sum_{j=d-r+2}^{d} \left( \frac{j^3 (j - 1) j (2j - 1)}{6} - \frac{j^2 (j - 1)^2 j^2}{4} \right) \]

\[ = \frac{2}{3} \sum_{j=d-r+2}^{d} (j^6 - j^4) \leq \frac{2}{3} \sum_{j=d-r+2}^{d} j^6 \]

\[ \leq \frac{2}{3} \cdot \frac{2^7}{7} \max\{(r - 2)^7, (r - 2)(d - r + 2)^6\} \]

\[ = O(\max(r^7, r(d - r)^6)) . \]

So we have obtained that

\[ E[\|S\|_{\text{H.S.}}] \leq \left( E[\|S\|_{\text{H.S.}}^2] \right)^{1/2} = O(\max(r^{7/2}, \sqrt{r}(d - r)^3)) . \]

As to \( E[\|T\|_{\text{H.S.}}] \), for \( n \geq 2d \) we have, again using Lemmas 3.2.9 and 3.2.13,

\[ E[\|T\|_{\text{H.S.}}^2] = \sum_{j,k=d-r+1}^{d} E[|t_{jk}|^2] = \sum_{j,k=d-r+1}^{d} 4j^2 k^2 E[p_{j+k}(M)p_{j+k}(M)] \]

\[ = \sum_{j,k=d-r+1}^{d} 4j^2 k^2 (j + k) = 4 \sum_{j,k=d-r+1}^{d} j^3 k^2 + 4 \sum_{j,k=d-r+1}^{d} k^3 j^2 \]

\[ = 8 \left( \sum_{j=d-r+1}^{d} j^3 \right) \left( \sum_{k=d-r+1}^{d} k^2 \right) \]

\[ \leq 8 \cdot \frac{2^4}{4} \max\{(r - 1)^4, (r - 1)(d - r + 1)^3\} \]

\[ \cdot \frac{2^3}{3} \max\{(r - 1)^3, (r - 1)(d - r + 1)^2\} \]

\[ = O(\max(r^7, r^2(d - r)^5)) . \]

Hence,
3. Multivariate normal approximation and the classical compact groups

\[ E[\|T\|_{\text{H.S.}}] \leq \left( E[\|T\|_{\text{H.S.}}^2] \right)^{1/2} = O(\max(r^{7/2}, (d-r)^{5/2})) . \]

Plugging these bounds into (3.42) yields

\[
d_W(W, Z_\Sigma) \leq \frac{1}{n(d-r+1)} \left( O(\max(r^3, \sqrt{r}(d-r)^{3/2})) \right.
\]
\[
+ \frac{O(\max(r^{7/2}, \sqrt{r}(d-r)^{3}) + \max(r^{7/2}, (d-r)^{5/2}))}{\sqrt{d-r+1}}
\]
\[
+ \frac{\max(r^{7/2}, \sqrt{r}(d-r)^{3}) + \max(r^{7/2}, (d-r)^{5/2})}{n(d-r+1)^{3/2}} \bigg) .
\]

Considering the cases \( r \leq d-r \) and \( d-r < r \) separately yields that

\[ d_W(W, Z_\Sigma) = O \left( \frac{\max \left\{ r^{7/2}, \sqrt{r}(d-r)^{3} \right\}}{n(d-r+1)^{3/2}} \right), \]

proving the first claim in Theorem 3.2.1. The others follow easily from that. Hence, the proof of Theorem 3.2.1 in the case \( K_n = U(n) \) is complete.

\[ \Box \]

**Proof of Theorem 3.2.1 in the special orthogonal case.** In this case we have to check the conditions (i)-(iii) of Theorem 3.1.9 for the family of exchangeable pairs \((W, W_t), t > 0, \) from Construction 3.2.3. In order to verify Condition (i), we will need the following lemma.

**Lemma 3.2.17.** For all \( j = d-r+1, \ldots, d \)

\[ E[W_t, j - W_j | M] = E[f_j(M_t) - f_j(M) | M] = t \cdot \left( -\frac{(n-1)j}{2} f_j(M) + R_j + O(t) \right), \]

where

\[
R_j = \begin{cases} 
-\frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) + \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) & \text{if } j \text{ is odd,} \\
-\frac{(n-1)j}{2} - \frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) + \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) & \text{if } j \text{ is even.}
\end{cases}
\]

142
3.2. Spectral properties of large random matrices from the classical compact groups

**Proof.** First observe that always \( f_j(M_t) - f_j(M) = p_j(M_t) - p_j(M) \), no matter what the parity of \( j \) is. By Lemmas 3.2.4 and 3.2.7

\[
E[p_j(M_t) - p_j(M) | M] = t(\Delta p_j)(M) + O(t^2)
\]

which is

\[
t \left( -\frac{(n-1)}{2} f_j(M) + \left( -\frac{1}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) + \frac{1}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \right) + O(t) \right)
\]

if \( j \) is odd and which is

\[
t \left( -\frac{(n-1)}{2} f_j(M) + \left( -\frac{1}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) + \frac{1}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \right) + O(t) \right)
\]

if \( j \) is even. This proves the lemma.

From Lemma 3.2.17 (see also Remark 3.2.5) we conclude

\[
\frac{1}{t} E[W_t - W | M] \xrightarrow{t \to 0} -\Lambda W + R \text{ almost surely and in } L^1(P),
\]

where \( \Lambda = \text{diag}(\frac{(n-1)}{2}, j = d-r+1, \ldots, d) \) and \( R = (R_{d-r+1}, \ldots, R_d)^T \). Thus, Condition (i) of Theorem 3.1.9 is satisfied. In order to verify Condition (ii) we will first prove the following

**Lemma 3.2.18.** For all \( j, k = d-r+1, \ldots, d \)

\[
E[(p_j(M_t) - p_j(M))(p_k(M_t) - p_k(M)) | M] = t(jkp_{j-k}(M) - jkp_{j+k}(M)) + O(t^2).
\]

**Proof.** By well-known properties of conditional expectation

\[
E[(p_j(M_t) - p_j(M))(p_k(M_t) - p_k(M)) | M] = E[p_j(M_t) | M] - p_j(M)E[p_k(M_t) | M] - p_k(M)E[p_j(M_t) | M] + p_{j+k}(M).
\]

(3.43)
3. Multivariate normal approximation and the classical compact groups

\[ E[p_{j,k}(M_t) \mid M] = p_{j,k}(M) + t(\Delta p_{j,k})(M) + O(t^2) \]

\[ = p_{j,k}(M) + t \left( -\frac{(n-1)(j+k)}{2} p_{j,k}(M) - \frac{j}{2} p_k(M) \sum_{l=1}^{j-1} p_{l,j-l}(M) \right) \]

\[ - \frac{k}{2} p_j(M) \sum_{l=1}^{k-1} p_{l,k-l}(M) - kjp_{j+k}(M) + \frac{j}{2} p_k(M) \sum_{l=1}^{j-1} p_{j-2l}(M) \]

\[ + \frac{k}{2} p_j(M) \sum_{l=1}^{k-1} p_{k-2l}(M) + jkp_{j-k}(M) \right) + O(t^2). \]

Also, by Lemma 3.2.4 and Claim (i) of Lemma 3.2.7

\[ - p_j(M)E[p_k(M_t) \mid M] = -p_j(M)(p_k(M) + t(\Delta p_k)(M) + O(t^2)) \]

\[ = -p_{j,k}(M) - tp_j(M) \left( -\frac{(n-1)k}{2} p_k(M) - \frac{k}{2} \sum_{l=1}^{k-1} p_{l,k-l}(M) + \frac{k}{2} \sum_{l=1}^{k-1} p_{2l-k}(M) \right) \]

\[ + O(t^2) \]

\[ = -p_{j,k}(M) + \frac{(n-1)k}{2} p_{j,k}(M) + \frac{k}{2} p_j(M) \sum_{l=1}^{k-1} p_{l,k-l}(M) \]

\[ - \frac{k}{2} p_j(M) \sum_{l=1}^{k-1} p_{2l-k}(M) + O(t^2), \]

and for reasons of symmetry

\[ -p_k(M)E[p_j(M_t) \mid M] = -p_{j,k}(M) + t \left( \frac{n-1}{2} p_{j,k}(M) + \frac{j}{2} p_k(M) \sum_{l=1}^{j-1} p_{l,j-l}(M) \right) \]

\[ - \frac{j}{2} p_k(M) \sum_{l=1}^{j-1} p_{2l-j}(M) + O(t^2). \]

Plugging these results into (3.43) above and noting that for \( j \in \mathbb{N} \)

\[ p_{-j}(M) = \text{Tr}(M^{-j}) = \text{Tr}((M^T)^j) = \text{Tr}(M^j)^T = \text{Tr}(M^j) = p_j(M), \]

we see that many terms cancel, and finally obtain

\[ E[(p_j(M_t) - p_j(M))(p_k(M_t) - p_k(M)) \mid M] = t(jkp_{j-k}(M) - jkp_{j+k}(M)) + O(t^2). \]
3.2. Spectral properties of large random matrices from the classical compact groups

Observing as above that, regardless of the parity of \( j \), we have \( f_j(M_t) - f_j(M) = p_j(M_t) - p_j(M) \), we can now easily compute

\[
\frac{1}{t} E[(W_{t,j} - W_j)(W_{t,k} - W_k) | M] = \frac{1}{t} E[(f_j(M_t) - f_j(M))(f_k(M_t) - f_k(M)) | M]
\]

\[
= \frac{1}{t} E[(p_j(M_t) - p_j(M))(p_k(M_t) - p_k(M)) | M]
\]

\[
= jk p_{j-k}(M) - jk p_{j+k}(M) + O(t)
\]

\[
t \to 0 \quad jk p_{j-k}(M) - jk p_{j+k}(M) \text{ a.s. and in } L^1(P),
\]

for all \( j, k = 1, \ldots, d \). Noting that for \( j = k \) the last expression is \( j^2 n - j^2 p_{2j}(M) \) and that \( 2 \Lambda \Sigma = \text{diag}((n-1)j^2, j = 1, \ldots, d) \) we see that Condition (ii) of Theorem 3.1.9 is satisfied with the matrix \( S = (S_{j,k})_{j,k=1,\ldots,d} \) given by

\[
S_{j,k} = \begin{cases} 
  j^2(1 - p_{2j}(M)), & j = k \\
  jk p_{j-k}(M) - jk p_{j+k}(M), & j \neq k.
\end{cases}
\]

In order to show that Condition (iii) of Theorem 3.1.9 holds, we will need the following facts:

**Lemma 3.2.19.** For all \( j = 1, \ldots, d \), \( n \geq 4d + 1 \),

\[
E[(W_{t,j} - W_j)^2] = t j^2(n - 1) + O(t^2). \quad \text{(i)}
\]

\[
E[(W_{t,j} - W_j)^4] = O(t^2). \quad \text{(ii)}
\]

**Proof.** As for (i), by Lemma 3.2.18,

\[
E[(W_{t,j} - W_j)^2] = E[(p_j(M_t) - p_j(M))^2]
\]

\[
= E[E[(p_j(M_t) - p_j(M))^2 | M]] = E[t j^2(n - p_{2j}(M)) + O(t^2)]
\]

\[
= t j^2(n - 1) + O(t^2),
\]

since by Lemma 3.2.10 \( E[p_{2j}(M)] = 1 \). For Claim (ii) we compute

\[
E[(W_{t,j} - W_j)^4] = E[(p_j(M_t) - p_j(M))^4]
\]

\[
= E[p_j(M_t)^4] + E[p_j(M)^4] - 4E[p_j(M_t)^3 p_j(M)] - 4E[p_j(M)^3 p_j(M_t)]
\]

\[
+ 6E[p_j(M_t)^2 p_j(M)^2]
\]

\[
= 2E[p_j(M)^4] - 8E[p_j(M)^3 p_j(M_t)] + 6E[p_{2j}(M)p_{j,j}(M_t)],
\]

145
3. Multivariate normal approximation and the classical compact groups

where the last equality follows from exchangeability. By Lemma 3.2.4 and Claim (ii) of Lemma 3.2.7 for the case $k = j$

$$E[p_{j,j}(M)p_{j,j}(M_t)] = E[p_{j,j}(M)E[p_{j,j}(M_t) | M]]$$
$$= E[p_{j,j}(M)(p_{j,j}(M) + t(\Delta p_{j,j})(M) + O(t^2))]$$
$$= E[p_j(M)^4] + tE[p_{j,j}(M) \cdot \Delta p_{j,j}(M)] + O(t^2)$$

$$= E[p_j(M)^4] + tE \left[ p_{j,j}(M) \left( -(n-1)jp_{j,j}(M) - jp_j(M) \sum_{l=1}^{j-1} p_{l,j-l}(M) - j^2p_{2j}(M) 
+ jp_j(M) \sum_{l=1}^{j-1} p_{j-2l}(M) + j^2n \right) \right] + O(t^2).$$

Again by Lemma 3.2.4 and Claim (i) of Lemma 3.2.7,

$$E[p_j(M)^3p_j(M_t)] = E[p_j(M)^3E[p_j(M_t) | M]]$$
$$= E[p_j(M)^3((p_j(M) + t(\Delta p_j)(M) + O(t^2))]$$
$$= E[p_j(M)^4] + tE \left[ p_j(M)^3 \left( -\frac{(n-1)}{2}p_j(M) - \frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) + \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \right) \right] + O(t^2).$$

Therefore,

$$E[(W_{l,j} - W_j)^4]$$

$$= 2E[p_j(M)^4] - 8 \left( E[p_j(M)^4] - t \frac{(n-1)}{2}E[p_j(M)^4] \right)$$
$$- t^j \frac{j}{2} \sum_{l=1}^{j-1} E[p_j(M)^3p_{l,j-l}(M)] + t^j \frac{j}{2} \sum_{l=1}^{j-1} E[p_j(M)^3p_{2l-j}(M)] \right)$$
$$+ 6 \left( E[p_j(M)^4] - t(n-1)jE[p_j(M)^4] - t \sum_{l=1}^{j-1} E[p_j(M)^3p_{l,j-l}(M)] \right)$$
$$- t^j E[p_{2j}(M)p_j(M)^2] + t^j \sum_{l=1}^{j-1} E[p_j(M)^3p_{2l-j}(M)] + t^j nE[p_j(M)^2] \right) + O(t^2)$$
$$= -2t(n-1)jE[p_j(M)^4] - 2t^j \sum_{l=1}^{j-1} E[p_j(M)^3p_{l,j-l}(M)] + 2t^j \sum_{l=1}^{j-1} E[p_j(M)^3p_{2l-j}(M)]$$
$$- 6t^j E[p_{2j}(M)p_j(M)^2] + 6t^j nE[p_j(M)^2] + O(t^2).$$

146
3.2. Spectral properties of large random matrices from the classical compact groups

Case 1: \( j \) is odd. Then by Lemma 3.2.10

\[
E[(W_{t,j} - W_j)^4] = -2t(n-1)j \cdot 3j^2 - 0 + 0 - 6tj^2 \cdot j + 6tj^2n \cdot j + O(t^2)
\]

\[= O(t^2), \]

as claimed.

Case 2: \( j \) is even. Then, again by Lemma 3.2.10

\[
E[(W_{t,j} - W_j)^4] = -2t(n-1)j \cdot (1 + 6j^2) - 2tj(1 + 3j) \sum_{l=1}^{j-1} E[p_{l,j-l}(M)]
\]

\[+ 2tj(1 + 3j) \sum_{l=1}^{j-1} E[p_{2l-j}(M)] - 6tj^2(1 + j) + 6tj^2n(1 + j) + O(t^2).
\]

Consider the term \( \sum_{l=1}^{j-1} E[p_{l,j-l}(M)] \). Since \( E[p_{l,j-l}(M)] = 0 \) whenever \( l \) is odd and \( l \neq j/2 \), we can write

\[
\sum_{l=1}^{j-1} E[p_{l,j-l}(M)] = E[p_{j/2,j/2}(M)] + \sum_{k=1}^{j/2-1} E[p_{2k,j-2k}(M)]
\]

\[= E[p_{j/2,j/2}(M)] + \sum_{k=1}^{j/2-1} E[p_{j/2,j/2}(M)] + \sum_{k=1}^{j/2-1} 1,
\]

where the last equality follows again by Lemma 3.2.10. If \( j/2 \) is odd, then, clearly, \( k \neq j/4 \) is not really a restriction, and by Lemma 3.2.10

\[
\sum_{l=1}^{j-1} E[p_{l,j-l}(M)] = \frac{j}{2} + \left( \frac{j}{2} - 1 \right) = j - 1.
\]

If \( j/2 \) is even, then by Lemma 3.2.10

\[
\sum_{l=1}^{j-1} E[p_{l,j-l}(M)] = \left(1 + \frac{j}{2}\right) + \left( \frac{j}{2} - 2 \right) = j - 1.
\]

Hence, in either case we have \( \sum_{l=1}^{j-1} E[p_{l,j-l}(M)] = j - 1 \).

Furthermore
3. Multivariate normal approximation and the classical compact groups

\[ \sum_{l=1}^{j-1} E[p_{2l-j}(M)] = 2 \sum_{l=1}^{j/2-1} E[p_{2l-j}(M)] + E[\text{Tr}(I_n)] = 2\left(\frac{j}{2} - 1\right) + n = j - 2 + n. \]

Therefore, if \( j \) is even

\[ E[(_W_t^j - W_j)^4] = -2t_j(n - 1)(1 + 6j + 3j^2) - 2t_j(1 + 3j)(j - 1) \]
\[ + 2t_j(1 + 3j)(j - 2 + n) - 6t_j^2(1 + j) + 6t_j^2n(1 + j) + O(t^2) \]
\[ = t_j(-2(n-1) + 2 + 2(n-2)) \]
\[ + t_j^2(-12(n-1) - 2 + 6 + 2(n-2) - 6 + 6n) \]
\[ + t_j^3(-6(n-1) - 6 + 6 - 6n) + O(t^2) \]
\[ = O(t^2), \]

as asserted. \( \square \)

Now we are in a position to check Condition (iii)' of Remark 3.1.10 and thereby prove Condition (iii) of Theorem 3.1.9. By Hölder’s inequality

\[ E[\|W_t - W\|_2^3] = E\left[\left( \sum_{j=d-r+1}^{d} (W_{t,j} - W_j)^2 \right)^{3/2} \right] \]
\[ = E\left[ \left( \sum_{j,k,l=d-r+1}^{d} (W_{t,j} - W_j)^2(W_{t,k} - W_k)^2(W_{t,l} - W_l)^2 \right)^{1/2} \right] \]
\[ \leq \sum_{j,k,l=d-r+1}^{d} E[|(W_{t,j} - W_j)(W_{t,k} - W_k)(W_{t,l} - W_l)|] \]
\[ \leq \sum_{j,k,l=d-r+1}^{d} \left( E[|W_{t,j} - W_j|^3] E[|W_{t,k} - W_k|^3] E[|W_{t,l} - W_l|^3] \right)^{1/3}. \]

Thus we have

\[ \frac{1}{t^5} E[\|W_t - W\|_2^3] \]
\[ \leq \sum_{j,k,l=d-r+1}^{d} \left( \frac{1}{t^5} E[|W_{t,j} - W_j|^3] \frac{1}{t^5} E[|W_{t,k} - W_k|^3] \frac{1}{t^5} E[|W_{t,l} - W_l|^3] \right)^{1/3}, \]
3.2. Spectral properties of large random matrices from the classical compact groups

and it suffices to show that for all \( j = d - r + 1, \ldots, d \)

\[
\lim_{t \to 0} \frac{1}{t} E \left[ |W_{t,j} - W_j|^3 \right] = 0.
\]

But this follows from Lemma 3.2.19, since by the Cauchy-Schwarz inequality

\[
E \left[ |W_{t,j} - W_j|^3 \right] \leq \sqrt{E \left[ |W_{t,j} - W_j|^2 \right] E \left[ |W_{t,j} - W_j|^4 \right]}
= \sqrt{(t^2(n - 1) + O(t^2)) \cdot O(t^2)}
= O(t^3/2)
\]

and hence \( \frac{1}{t} E \left[ |W_{t,j} - W_j|^3 \right] \xrightarrow{t \to 0} 0. \)

By Theorem 3.1.9 we can now conclude that

\[
d_{W}(W, Z, \Sigma) \leq \|\Lambda^{-1}\|_{op} \left( E[\|R\|_2^2] + \frac{1}{\sqrt{2\pi}} \|\Sigma^{-1/2}\|_{op} E[\|S\|_{H.S.}] \right).
\]

Clearly, \( \|\Lambda^{-1}\|_{op} = \frac{2}{(n-1)(d-r+1)} \) and \( \|\Sigma^{-1/2}\|_{op} = \frac{1}{\sqrt{d-r+1}} \). In order to bound \( E[\|R\|_2^2] \) we will first prove the following lemma.

**Lemma 3.2.20.** If \( n \geq 4d + 1 \), for all \( j = d - r + 1, \ldots, d \) we have that \( E[R_j^2] = O(j^5) \).

**Proof.** First suppose that \( j \) is odd. Then

\[
E[R_j^2] = E \left[ \left( -\frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) + \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \right)^2 \right]
= \frac{j^2}{4} \left( \sum_{l,k=1}^{j-1} E[p_{l,j-l}(M)p_{k,j-k}(M)] - 2 \sum_{l,k=1}^{j-1} E[p_{l,j-l}(M)p_{2k-j}(M)] 
+ \sum_{l,k=1}^{j-1} E[p_{2l-j}(M)p_{2k-j}(M)] \right) = \frac{j^2}{4} (T_1 - 2T_2 + T_3).
\]

By Lemma 3.2.10 for \( j \) odd
3. Multivariate normal approximation and the classical compact groups

\[ E[p_{i,j-l}(M)p_{k,j-k}(M)] = \begin{cases} 
  l(j-l+1), & \text{if } k = l \text{ is odd}, \\
  (l+1)(j-l), & \text{if } k = l \text{ is even}, \\
  l(j-l+1), & \text{if } l \text{ is odd and } k = j - l, \\
  0, & \text{if } l \text{ is odd and } k \notin \{l, j-l\}, \\
  (l+1)(j-l), & \text{if } l \text{ is even and } k = j - l, \\
  0, & \text{if } l \text{ is even and } k \notin \{l, j-l\}, 
\end{cases} \]

\[ E[p_{i,j-l}(M)p_{2k-j}(M)] = \begin{cases} 
  l, & \text{if } l = 2k - j, \\
  j-l, & \text{if } l = 2j - 2k, \\
  l, & \text{if } l = j - 2k, \\
  j-l, & \text{if } l = 2k, \\
  0, & \text{otherwise,} 
\end{cases} \]

\[ E[p_{2i-j}(M)p_{2k-j}(M)] = \begin{cases} 
  |2l-j|, & \text{if } l = k \text{ or } l = j - k, \\
  0, & \text{otherwise.} 
\end{cases} \]

Therefore,

\[ T_1 = 2 \sum_{l=1}^{j-1} l(j-l+1) + 2 \sum_{l=1}^{j-1} (l+1)(j-l) \]

\[ \leq 2j \sum_{l=1}^{j} l = j^2(j-1) = O(j^3), \]

\[ T_2 = 2 \sum_{l=1}^{j-1} l + 2 \sum_{l=1}^{j-1} (j-l) \leq 2 \sum_{l=1}^{j} l = j(j-1) = O(j^2) \]

and

\[ T_3 = 2 \sum_{l=1}^{j-1} |2l-j| \leq 2 \sum_{l=1}^{j} j = 2j^2 = O(j^2). \]

Hence for \( j \) odd

\[ E[R_j^2] = \frac{j^2}{4} \left( O(j^3) - 2O(j^2) + O(j^2) \right) = O(j^5). \]
3.2. Spectral properties of large random matrices from the classical compact groups

The case that \( j \) is even can be treated similarly as can be seen by observing that in this case

\[
R_j = \frac{j}{2} \left( 1 - \sum_{l=1}^{j-1} p_{l,j-M} + \sum_{l=1}^{j-1} p_{2l-j} \right).
\]

But as in Case 2 of the proof of Lemma 3.2.19 (ii), one has to distinguish between the cases that \( j \equiv 0 \mod 4 \) or else that \( j \equiv 2 \mod 4 \), and the explicit formulae for \( E[R_j^2] \) are more complicated.

By Lemma 3.2.20 there is a constant \( C > 0 \) neither depending on \( j \) nor on \( n \) such that \( E[R_j^2] \leq Cj^5 \). Thus by Jensen’s inequality and Lemma 3.2.13 we obtain

\[
E[\|R\|_2] \leq \sqrt{\sum_{j=d-r+1}^{d} E[R_j^2]} \leq \sqrt{C} \sqrt{\sum_{j=d-r+1}^{d} j^5} \\
\leq \sqrt{C} \sqrt{\frac{2^6}{6} \max\{(r-1)^6, (r-1)(d-r+1)^5\}} \\
= O\left(\max\{r^3, (d-r)^{5/2}\} \right). \tag{3.44}
\]

Next, we turn to bounding \( E[\|S\|_{H.S.}] \). By Lemma 3.2.10 we have

\[
E[\|S\|_{H.S.}^2] = \sum_{j,k=d-r+1}^{d} E[S_{j,k}^2] = \sum_{j=d-r+1}^{d} j^4 E\left[1 - 2p_{2j}(M) + p_{2j}(M)^2\right] \\
+ \sum_{j,k=d-r+1}^{d} j^2k^2 E[p_{j-k}(M)^2 - 2p_{j-k}(M)p_{j+k}(M) + p_{j+k}(M)^2] \\
= 2 \sum_{j=d-r+1}^{d} j^5 + 2 \sum_{d-r+1 \leq k < j \leq d} j^2k^2 \left( E[p_{j-k}(M)^2] - 2E[p_{j-k}(M)p_{j+k}(M)] + E[p_{j+k}(M)^2]\right).
\]

Again, by Lemma 3.2.10
3. Multivariate normal approximation and the classical compact groups

\[ E[p_{j-k}(M)^2] = \begin{cases} 1 + j - k, & \text{if } j + k \text{ even}, \\ j - k, & \text{if } j + k \text{ odd}, \end{cases} \]

\[ E[p_{j-k}(M)p_{j+k}(M)] = \begin{cases} 1, & \text{if } j + k \text{ even}, \\ 0, & \text{if } j + k \text{ odd}, \end{cases} \]

\[ E[p_{j+k}(M)^2] = \begin{cases} 1 + j + k, & \text{if } j + k \text{ even} \\ j + k, & \text{if } j + k \text{ odd}. \end{cases} \]

Hence,

\[ E[\|S\|_{H.S.}^2] = 2 \sum_{j=d-r+1}^{d} j^5 + 2 \sum_{d-r+1 \leq k < j \leq d} j^2 k^2 (1 + j - k - 2 + 1 + j + k) \]

\[ + 2 \sum_{d-r+1 \leq k < j \leq d} j^2 k^2 (j - k - 2 \cdot 0 + j + k) \]

\[ = 2 \sum_{j=d-r+1}^{d} j^5 + 4 \sum_{d-r+1 \leq k < j \leq d} k^2 j^3. \]

From (3.44) we know that

\[ \sum_{j=d-r+1}^{d} j^5 = O\left( \max\{r^6, r(d-r)^5\} \right) \]

and furthermore we can bound

\[ \sum_{d-r+1 \leq k < j \leq d} k^2 j^3 \leq \sum_{j=d-r+2}^{d} j^2 (j - (d - r + 1)) j^3 \]

\[ = \sum_{j=2}^{r} (j + d - r)^2 (j + d - r - (d - r + 1)) (j + d - r)^3 \]

\[ = \sum_{j=2}^{r} (j + d - r)^5 (j - 1) \]

\[ \leq r \sum_{j=1}^{r} (j + d - r)^5 \leq 32r \max\{r^6, (d-r)^5r^2\} \]

\[ = 32 \max\{r^7, (d-r)^5r^2\}. \]
Thus, we obtain
\[
E[\|S\|_{\text{H.S.}}^2] = O\left(\max\{r^6, (d - r)^5 r\}\right) + 128 \max\{r^7, (d - r)^5 r^2\}
\]
and hence, by Jensen’s inequality
\[
E[\|S\|_{\text{H.S.}}] \leq \sqrt{E[\|S\|_{\text{H.S.}}^2]} = O\left(\max\{r^{7/2}, (d - r)^{5/2} r\}\right).
\]
Collecting terms, we see that
\[
d_{W}(W, Z_{\Sigma}) \leq \|\Lambda^{-1}\|_{\text{op}} E[\|R\|_{2}] + \frac{1}{\sqrt{2\pi}} \|\Lambda^{-1}\|_{\text{op}} \|\Sigma^{-1/2}\|_{\text{op}} E[\|S\|_{\text{H.S.}}]
\]
\[
= \frac{2}{(n - 1)(d - r + 1)} O\left(\max\{r^3, (d - r)^{5/2} \sqrt{r}\}\right)
\]
\[
+ \frac{1}{\sqrt{2\pi}} \frac{2}{(n - 1)(d - r + 1)^{3/2}} O\left(\max\{r^{7/2}, (d - r)^{5/2} r\}\right)
\]
\[
= O\left(\max\left\{\frac{r^3}{d - r + 1}, \frac{r^{7/2}}{(d - r + 1)^{3/2}}, \frac{(d - r)^{5/2} \sqrt{r}}{d - r + 1}, \frac{(d - r)^{5/2} r}{(d - r + 1)^{3/2}}\right\}\right).
\]
Proceeding as in the unitary case and treating the cases that \(r \leq d - r\) and \(d - r < r\) separately, one obtains that the last expression is of order
\[
\frac{1}{n} \max\left\{\frac{r^{7/2}}{(d - r + 1)^{3/2}}, (d - r)^{3/2} \sqrt{r}\right\}.
\]
This concludes the proof of Theorem 3.2.1 in the special orthogonal case.

Proof of Theorem 3.2.1 in the unitary symplectic case. In this case we have to check the conditions (i)-(iii) of Theorem 3.1.9 for the family of exchangeable pairs \((W, W_t)\), \(t > 0\), from Construction 3.2.3. In order to verify Condition (i), we will need the following lemma.

Lemma 3.2.21. For all \(j = d - r + 1, \ldots, d\)
\[
E[W_{t,j} - W_j \mid M] = E[f_j(M_t) - f_j(M) \mid M] = t \cdot \left(\frac{(n - 1)j}{2} f_j(M) + R_j + O(t)\right),
\]
where
3. Multivariate normal approximation and the classical compact groups

\[ R_j = -\frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) - \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \text{ if } j \text{ is odd,} \]

\[ R_j = \frac{(2n+1)j}{2} - \frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) - \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \text{ if } j \text{ is even.} \]

**Proof.** First observe that always \( f_j(M_t) - f_j(M) = p_j(M_t) - p_j(M) \), no matter what the parity of \( j \) is. By Lemmas 3.2.4 and 3.2.8

\[ E[p_j(M_t) - p_j(M) | M] = t(\Delta p_j)(M) + O(t^2) \]

\[ = t \left( -\frac{(2n+1)j}{2} p_j(M) - \frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) - \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \right) + O(t^2), \]

which is

\[ t \left( -\frac{(2n+1)j}{2} f_j(M) + \left( -\frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) - \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \right) + O(t) \right) \]

if \( j \) is odd and which is

\[ t \left( -\frac{(2n+1)j}{2} f_j(M) + \left( \frac{(2n+1)j}{2} - \frac{j}{2} \sum_{l=1}^{j-1} p_{l,j-l}(M) - \frac{j}{2} \sum_{l=1}^{j-1} p_{2l-j}(M) \right) + O(t) \right) \]

if \( j \) is even. This proves the lemma. \( \square \)

From Lemma 3.2.17 (see also Remark 3.2.5) we conclude

\[ \frac{1}{t} E[W_t - W | M] \xrightarrow{t \to 0} -\Lambda W + R \text{ a.s. and in } L^1(P), \]

where \( \Lambda = \text{diag}(\frac{(2n+1)j}{2}, j = d - r + 1, \ldots, d) \) and \( R = (R_{d-r+1}, \ldots, R_d)^T \). Thus, Condition (i) of Theorem 3.1.9 is satisfied. The validity of Condition (ii) follows from the next lemma, whose proof only differs from that of Lemma 3.2.18 in that it makes use of Lemma 3.2.8 in the place of Lemma 3.2.7.

**Lemma 3.2.22.** For all \( j, k = d - r + 1, \ldots, d, \)

\[ E[(p_j(M_t) - p_j(M))(p_k(M_t) - p_k(M)) | M] = t(jkp_{j-k}(M) - jkp_{j+k}(M)) + O(t^2). \]
3.2. Spectral properties of large random matrices from the classical compact groups

Observing that for all \( j = d - r + 1, \ldots, d \) we have \( f_j(M_t) - f_j(M) = p_j(M_t) - p_j(M) \) and using Lemma 3.2.22 we can now easily compute

\[
\frac{1}{t} E[(W_{t,j} - W_j)(W_{t,k} - W_k) | M] = \frac{1}{t} E[(f_j(M_t) - f_j(M))(f_k(M_t) - f_k(M)) | M]
\]

\[
= \frac{1}{t} E[(p_j(M_t) - p_j(M))(p_k(M_t) - p_k(M)) | M] = jk p_j - k(M) - jk p_j + k(M) + O(t) \quad t \to 0
\]

\[
jk p_j - k(M) - jk p_j + k(M) \text{ a.s. and in } L^1(P),
\]

for all \( d - r + 1 \leq j, k \leq d \). Noting that for \( j = k \) the last expression is \( 2 j^2 n - j^2 p_2 j(M) \) and that \( 2 \Lambda \Sigma = \text{diag}(2(n+1)j^2, j = d - r + 1, \ldots, d) \), we see that Condition (ii) of Theorem 3.1.9 is satisfied with the matrix \( S = (S_{j,k})_{j,k=d-r+1,\ldots,d} \) given by

\[
S_{j,k} = \begin{cases} 
-j^2(1 + p_2 j(M)), & j = k \\
jk p_j - k(M) - jk p_j + k(M), & j \neq k.
\end{cases}
\]

The validity of Condition (iii)’ of Remark 3.1.10 is based on the following lemma, which can be proven in the same way as its analog for the special orthogonal group, Lemma 3.2.19.

**Lemma 3.2.23.** If \( n \geq 2d \), for all \( j = 1, \ldots, d \) there holds

\[
E[(W_{t,j} - W_j)^2] = tj^2(2n + 1) + O(t^2). 
\]

\[
E[(W_{t,j} - W_j)^4] = O(t^2). 
\]

So we obtain from Theorem 3.1.9 that

\[
d_{W}(W, Z_{\Sigma}) \leq \|\Lambda^{-1}\|_{op} \left( E[\|R\|_2^2] + \frac{1}{\sqrt{2\pi}} \|\Sigma_{-1/2}\|_{op} E[\|S\|_{H.S.}] \right).
\]

Again, it is easy to see that \( \|\Lambda^{-1}\|_{op} = \frac{2}{(2n+1)(d-r+1)} \) and \( \|\Sigma_{-1/2}\|_{op} = \frac{1}{\sqrt{2\pi}} \). Since we can show in a similar way as in the proof for the special orthogonal case that \( E[\|R\|_2] = O(\max\{r^3, (d-r)^{5/2}\sqrt{r} \}) \) and

\[
E[\|S\|_{H.S.}] = O(\max\{r^{7/2}, (d-r)^{5/2}\sqrt{r} \}),
\]

we may conclude the proof of the unitary symplectic case as before.

\[ \square \]
3. Multivariate normal approximation and the classical compact groups

3.2.2. Linear eigenvalue statistics

In the present subsection we deal with the empirical measure of the eigenvalues of a Haar distributed random unitary matrix $M = M_n$. This random probability measure $L_n$ is defined by

$$L_n := \frac{1}{n} \sum_{j=1}^{n} \lambda_j ,$$

where $\lambda_1, \ldots, \lambda_n$ are the (exchangeable) eigenvalues of $M_n$. Since $M_n$ is unitary, they take values on the unit circle line $S^1$. In the following, let $\mu$ be the uniform distribution on the 1-sphere $S^1$, i.e. for each bounded and Borel measurable function $f : S^1 \to \mathbb{R}$ we have

$$\int_{S^1} f(z) \, d\mu(z) = \frac{1}{2\pi} \int_{0}^{2\pi} f(e^{it}) \, dt .$$

It was shown by Weyl, that the joint distribution of $\lambda_1, \ldots, \lambda_n$ has density

$$p(z_1, \ldots, z_n) := \frac{1}{n!} \prod_{1 \leq j < k \leq n} |z_k - z_j|^2$$

with respect to $\mu^{\otimes n}$. We will be concerned with asymptotic properties of the random measure $L_n$ as $n \to \infty$. It was already shown by Diaconis and Shahshahani [DS94], that as $n \to \infty$,

$$L_n \xrightarrow{w} \mu \text{ in probability},$$

where $w$ stands for weak convergence of probability measures. In the following, we will focus on the fluctuations of linear statistics of $L_n$ given by certain test functions $f$ on $S^1$. Note that for a fixed bounded and measurable function $f$ on $S^1$ the mapping $\nu \mapsto \int_{S^1} f(z) \, d\nu(z)$ is linear on the space of finite signed measures on $S^1$. It is a well known fact from calculus that if two arrays $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ of complex numbers are given such that

$$\sum_{j=1}^{n} a_j^k = \sum_{j=1}^{n} b_j^k \quad \text{for each integer } k \geq 1 ,$$

then there is a permutation $\sigma$ of $\{1, \ldots, n\}$ such that $b_j = a_{\sigma(j)}$ for each $1 \leq j \leq n$. Since $\text{Tr}(M_n^k) = \sum_{j=1}^{n} \lambda_j^k$, this simple fact implies that the sequence of traces of powers $\text{Tr}(M_n^k)$, $k \geq 1$, in principle carries the same information as the array $\lambda_1, \ldots, \lambda_n$ of eigenvalues of $M_n$. This equality of information was used in an explicit way by Diaconis and Evans [DE01], who proved Gaussian fluctuations for suitable statistics $L_n(f) = \int_{S^1} f(z) \, dL_n(z)$, by using the Fourier expansion of $f$ and
a central limit theorem for (infinite) $\mathbb{R}$-linear functionals of the sequence of traces of powers. It should be noted that it is already well known, that the fluctuations

$$n \left( L_n(f) - E[L_n(f)] \right)$$

are Gaussian for suitable functions $f$. In addition to [DE01], also Soshnikov [Sos00] proves such results and Johansson [Joh97] even obtains a super exponential rate of convergence but only for $f$ being a trigonometric polynomial. The aim of this subsection is to prove a rate of convergence for the Gaussian fluctuations of $L_n(f)$ for more general test functions $f$ than trigonometric polynomials. The crucial ingredients of our proof will be Theorem 3.2.1, particularly the fact that $d$ may be growing in $n$, and bounds on the Fourier coefficients of smooth test functions on $S^1$. For reasons of structure, we begin with some more or less easy auxiliary results. In this subsection we will write $\parallel h \parallel_{Lip} := M_1(h)$ for the Lipschitz seminorm introduced in Section 3.1.

**Lemma 3.2.24.** Let $\varphi : \mathbb{C}^r \to \mathbb{C}$. Then, the following properties are equivalent:

(a) $\varphi$ is $\mathbb{R}$-linear.

(b) There exist complex numbers $a_1, \ldots, a_r, b_1, \ldots, b_r$ such that for all $z = (z_1, \ldots, z_r)^T \in \mathbb{C}^r$ we have $\varphi(z) = \sum_{j=1}^r (a_j z_j + b_j \overline{z_j})$.

If this is the case, then one necessarily has $a_j = \frac{\varphi(e_j) - i\varphi(i e_j)}{2}$ and $b_j = \frac{\varphi(e_j) + i\varphi(i e_j)}{2}$ for $j = 1, \ldots, r$. In particular, this representation is unique and will be called canonical in what follows.

**Lemma 3.2.25.** Let $\varphi : \mathbb{C}^r \to \mathbb{C}$ be $\mathbb{R}$-linear with canonical representation $\varphi(z) = \sum_{j=1}^r (a_j z_j + b_j \overline{z_j})$. Then, it holds that $\parallel \varphi \parallel_{Lip} = \parallel \varphi \parallel_{op} \leq \left( \parallel a \parallel_2 + \parallel b \parallel_2 \right)$, where $a = (a_1, \ldots, a_r)^T$ and $b = (b_1, \ldots, b_r)^T$.

We will need the following special versions of the Diaconis-Shahshahani formulæ, see also Lemma 3.2.9.

**Lemma 3.2.26.** Let $n, j, k$ be positive integers. Then,

$$E[\text{Tr}(M_n^j)] = 0,$$

$$E[\text{Tr}(M_n^j) \text{Tr}(M_n^k)] = 0 \quad \text{and}$$

$$E\left[ \frac{\text{Tr}(M_n^j)}{\text{Tr}(M_n^k)} \right] = \delta_{jk} (j \wedge n).$$
3. Multivariate normal approximation and the classical compact groups

Proof. The first two claims are immediately implied by Lemma 3.2.9 as is the third one in the case that either $j \neq k$ or $j = k$ and $n \geq j$. If $j = k$ and $n < j$, then by Rains’ theorem [Rai97b], the eigenvalues $\lambda^1_j, \ldots, \lambda^n_j$ are independent and uniformly distributed on $S^1$. Since the uniform distribution on $S^1$ has mean zero, this implies that

$$E\left[\text{Tr}(M^n_j)\text{Tr}(M^n_j)\right] = \sum_{l,m=1}^{n} E[\lambda^l_j \lambda^m_j] = \sum_{l=1}^{n} E[|\lambda^l_j|^2] = n.$$ 

We will need some facts from the theory of Fourier series. Consider a $C$-valued function $f \in L^2(S^1, B(S^1), \mu)$, where $B(S^1)$ is the $\sigma$-field of Borel subsets of $S^1$. By the definition $\hat{f}(x) := f(e^{ix})$ we obtain a $2\pi$-periodic locally integrable function $\hat{f} : \mathbb{R} \to \mathbb{R}$. In the following, we will tacitly identify $f$ and $\hat{f}$. Recall that for $j \in \mathbb{Z}$ the $j$-th Fourier coefficient of $f$ is defined as

$$\hat{f}(j) := \int_{S^1} f(z) z^{-j} d\mu(z) = \frac{1}{2\pi} \int_0^{2\pi} \hat{f}(e^{ix}) e^{-ijx} dx.$$ 

Thus, if $f$ is real-valued, then $\hat{f}(-j) = \overline{\hat{f}(j)}$ for each $j \in \mathbb{Z}$. Henceforth, we will always assume that $f$ is a real-valued function on $S^1$. For our purposes, the crucial fact about Fourier coefficients is that smoothness of the function $f$ implies quantitative information on the decay of Fourier coefficients. Thus, in the following, we will only deal with $C^k$-functions. Note that $f$ and $\hat{f}$ have the same smoothness properties.

**Proposition 3.2.27.** Let $k \geq 1$ be a positive integer. If $f \in C^k(S^1)$, then for each $j \in \mathbb{Z} \setminus \{0\}$ it holds that

$$|\hat{f}(j)| \leq \frac{\|f^{(k)}\|_1}{|j|^k}.$$ 

Proof. See, e.g. [Kat04, I.4.4].

It is known (see, e.g. [Kat04, II.2.2]) that if $f$ is continuous and of bounded variation (for example $f \in C^k(S^1)$, $k \geq 1$) then it is the uniform and, hence, pointwise limit of its Fourier series

$$\sum_{j \in \mathbb{Z}} \hat{f}(j) e^{ijx} = \hat{f}(0) + \sum_{j=1}^{\infty} \left( \hat{f}(j) e^{ijx} + \overline{\hat{f}(j)} e^{-ijx} \right),$$

where the last identity is because $f$ is real-valued. Thus, we can write
3.2. Spectral properties of large random matrices from the classical compact groups

\[ L_n(f) = \frac{1}{n} \sum_{k=1}^{n} f(\lambda_k) = \frac{1}{n} \sum_{k=1}^{n} \left( \hat{f}(0) + \sum_{j=1}^{\infty} \left( \hat{f}(j) \lambda_k^j + \overline{\hat{f}(j) \lambda_k^j} \right) \right) \]

\[ = \hat{f}(0) + \frac{1}{n} \sum_{j=1}^{\infty} \left( \hat{f}(j) \sum_{k=1}^{n} \lambda_k^j + \overline{\hat{f}(j) \sum_{k=1}^{n} \lambda_k^j} \right) \]

\[ = \hat{f}(0) + \frac{1}{n} \sum_{j=1}^{\infty} \left( \hat{f}(j) \text{Tr}(M_n^j) + \overline{\hat{f}(j) \text{Tr}(M_n^j)} \right). \quad (3.45) \]

Using (3.45) and Lemma 3.2.26, one can easily show that \( E[L_n(f)] = \hat{f}(0) \) and, thus we obtain the following representation for the \( n \)-scaled fluctuation of \( L_n(f) \):

\[ n(L_n(f) - E[L_n(f)]) = \sum_{j=1}^{\infty} \left( \hat{f}(j) \text{Tr}(M_n^j) + \overline{\hat{f}(j) \text{Tr}(M_n^j)} \right) \quad (3.46) \]

Formula (3.46) shows explicitly how the sequence of traces of powers contains the whole information of the eigenvalues \( \lambda_1, \ldots, \lambda_n \) which is important to compute \( L_n(f) \). This representation is one of several examples that motivated Diaconis and Evans [DE01] to study, more generally, \( \mathbb{R} \)-linear functionals of the sequence of traces of powers, i.e. expressions of the form

\[ \sum_{j=1}^{\infty} \left( a_{n,j} \text{Tr}(M_n^j) + b_{n,j} \overline{\text{Tr}(M_n^j)} \right) \]

for suitable double arrays \((a_{n,j})_{n,j \in \mathbb{N}}\) and \((b_{n,j})_{n,j \in \mathbb{N}}\) of complex numbers.

Before stating the main theorem of this subsection, let us fix some notation. Throughout let \( Z = (Z_1, \ldots, Z_d)^T \sim N_{d, \mathbb{C}}(0, I_d) \) be a \( d \)-dimensional standard complex normal random vector, let \( \Sigma := \text{diag}(1, 2, \ldots, d) \), \( Z_{\Sigma} := \Sigma^{1/2} Z \sim N_{d, \mathbb{C}}(0, \Sigma) \) and let \( \zeta \sim N(0, 1) \) be a real one-dimensional standard normal random variable.

**Theorem 3.2.28.** Let \( M_n \) be chosen from Haar measure on the unitary group \( U(n) \) and let \( f : S^1 \to \mathbb{R} \) be a \( C^k \)-function for some integer \( k \geq 2 \). Let \( \sigma > 0 \) be defined by

\[ \sigma^2 := \sum_{j \in \mathbb{Z}} |\hat{f}(j)|^2 |j| = 2 \sum_{j=1}^{\infty} j |\hat{f}(j)|^2 < \infty. \]

Then, with \( S := S_n := n(L_n(f) - E[L_n(f)]) \) the Wasserstein distance between \( \mathcal{L}(S) \) and \( N(0, \sigma^2) \) can be bounded by

159
3. Multivariate normal approximation and the classical compact groups

\[ d_{W}(\mathcal{L}(S), N(0, \sigma^2)) \leq C n^{-\left( \frac{k-1}{k+5/2} \right)}, \]

where \( C > 0 \) is a finite constant, which only depends on \( f \) and \( k \). In particular, if \( f \) is \( C^\infty \), then for each \( \varepsilon > 0 \) we have

\[ d_{W}(\mathcal{L}(S), N(0, \sigma^2)) = O\left( \frac{1}{n^{1-\varepsilon}} \right). \]

**Proof.** Without loss of generality we may assume that \( \sigma^2 > 0 \), because otherwise \( f = 0 \) and the assertion is trivially true. It follows from Proposition 3.2.27 that \( \sigma^2 < \infty \). By (3.46) we have

\[ S = S_n = \sum_{j=1}^{\infty} \left( \hat{f}(j) \text{Tr}(M_n^j) + \overline{\hat{f}(j)} \text{Tr}(M_n^j) \right). \]

For \( d = d(n) \leq \frac{n}{2} \) to be chosen later, write

\[ S^{(1)} = S_n^{(1)} = \sum_{j=1}^{d} \left( \hat{f}(j) \text{Tr}(M_n^j) + \overline{\hat{f}(j)} \text{Tr}(M_n^j) \right) \]

and \( S^{(2)} = S_n^{(2)} = S - S^{(1)} \). Then, by Lemma 3.2.26, Proposition 3.2.27 and the dominated convergence theorem,

\[
E\left[ |S^{(2)}|^2 \right] = E \left[ \sum_{j,l=d+1}^{\infty} \left( \hat{f}(j) \text{Tr}(M_n^j) + \overline{\hat{f}(j)} \text{Tr}(M_n^j) \right) \left( \hat{f}(l) \text{Tr}(M_n^l) + \overline{\hat{f}(l)} \text{Tr}(M_n^l) \right) \right] \\
= \sum_{j,l=d+1}^{\infty} E \left[ \left( \hat{f}(j) \text{Tr}(M_n^j) + \overline{\hat{f}(j)} \text{Tr}(M_n^j) \right) \left( \hat{f}(l) \text{Tr}(M_n^l) + \overline{\hat{f}(l)} \text{Tr}(M_n^l) \right) \right] \\
= \sum_{j,l=d+1}^{\infty} \left( \hat{f}(j) \overline{\hat{f}(l)} E \left[ \text{Tr}(M_n^j) \text{Tr}(M_n^l) \right] + \hat{f}(j) \hat{f}(l) E \left[ \text{Tr}(M_n^j) \text{Tr}(M_n^l) \right] \right) \\
+ \hat{f}(j) \overline{\hat{f}(l)} E \left[ \text{Tr}(M_n^j) \text{Tr}(M_n^l) \right] + \overline{\hat{f}(j)} \overline{\hat{f}(l)} E \left[ \text{Tr}(M_n^j) \text{Tr}(M_n^l) \right] \\
= 2 \sum_{j=d+1}^{\infty} |\hat{f}(j)|^2 (j \wedge n).
\]

Hence, by Proposition 3.2.27 we obtain
3.2. Spectral properties of large random matrices from the classical compact groups

\[ E[|S^{(2)}|^2] \leq 2 \sum_{j=1}^{\infty} j |\hat{f}(j)|^2 \leq 2 \|f^{(k)}\|_1^2 \sum_{j=1}^{\infty} \frac{1}{j^{2k-1}} \]

\[ \leq 2 \|f^{(k)}\|_1^2 \int_1^\infty \frac{1}{x^{2k-1}} dx \]

\[ = \frac{2\|f^{(k)}\|_1^2}{(2k-2)d^{2k-2}} \]

\[ = \frac{\|f^{(k)}\|_1^2}{(k-1)d^{2k-2}}. \] (3.47)

Define \( \varphi : \mathbb{C}^d \to \mathbb{R} \) by

\[ \varphi(z) := \sum_{j=1}^{d} (\hat{f}(j)z_j + \overline{\hat{f}(j)}z_j). \]

By Lemma 3.2.24 \( \varphi \) is \( \mathbb{R} \)-linear and with \( a := ((\hat{f}(1), \ldots, (\hat{f}(d))^T \) and \( b := \overline{a} \) by Lemma 3.2.25 we have

\[ \|\varphi\|_{\text{Lip}} \leq \|a\|_2 + \|b\|_2 = 2\|a\|_2. \]

Note that by Proposition 3.2.27

\[ \|a\|_2^2 = \sum_{j=1}^{d} |\hat{f}(j)|^2 \leq \|f^{(k)}\|_1^2 \sum_{j=1}^{\infty} \frac{1}{j^{2k}} < \infty \]

is bounded by a constant, which is independent of \( n \) and of the specific form of \( f \).

Now, let \( g : \mathbb{R} \to \mathbb{R} \) be Lipschitz continuous with \( \|g\|_{\text{Lip}} \leq 1 \). Then,

\[ \left| E[g(S)] - E[g(\sigma\zeta)] \right| \leq \left| E[g(S)] - E[g(S^{(1)})] \right| + \left| E[g(S^{(1)})] - E[g(\varphi(Z\Sigma))] \right| \]

\[ + \left| E[g(\varphi(Z\Sigma))] - E[g(\sigma\zeta)] \right| \] (3.48)

Using (3.47) we can bound the first summand in (3.48) as follows:

\[ \left| E[g(S)] - E[g(S^{(1)})] \right| \leq \|g\|_{\text{Lip}} E[|S - S^{(1)}|] \leq E[|S^{(2)}|] \]

\[ \leq \left( E[|S^{(2)}|^2] \right)^{1/2} \leq \frac{\|f^{(k)}\|_1}{\sqrt{k-1}d^{k-1}}. \] (3.49)
Writing \( W := (\text{Tr}(M_n), \ldots, \text{Tr}(M_n^d))^T \), we see that \( S^{(1)} = \varphi(W) \). Since

\[
\|g \circ \varphi\|_{Lip} \leq \|g\|_{Lip} \|\varphi\|_{Lip} \leq 2\|a\|_2,
\]

we can conclude from Theorem 3.2.1 with \( r = d \) for the second summand in (3.48) that

\[
\left| E[g(S^{(1)})] - E[g(\varphi(Z_\Sigma))] \right| = \left| E[g(\varphi(W))] - E[g(\varphi(Z_\Sigma))] \right| \leq 2\|a\|_2 d_{W}(W, Z_\Sigma) \leq 2K\|a\|_2 \frac{d^{7/2}}{n},
\]

where \( K > 0 \) is the constant from Theorem 3.2.1. To bound the third summand in (3.48), we will first study the distribution of \( \varphi(Z_\Sigma) \). Recall that \( Z_\Sigma \) is the random vector \((Z_1, \sqrt{2}Z_2, \ldots, \sqrt{d}Z_d)^T\), where \( Z_1, \ldots, Z_d \) are iid standard complex normal random variables, i.e. there are iid real normal random variables \( X_1, \ldots, X_d, Y_1, \ldots, Y_d \) with distribution \( N(0,1/2) \) such that \( Z_j = X_j + iY_j \) for \( 1 \leq j \leq d \). Thus we obtain

\[
\varphi(Z_\Sigma) = \sum_{j=1}^{d} \left( \hat{f}(j)\sqrt{j}Z_j + \overline{\hat{f}(j)}\sqrt{j}Z_j \right)
\]

\[
= 2 \sum_{j=1}^{d} \sqrt{j} \text{Re}(\hat{f}(j)Z_j)
\]

\[
= 2 \sum_{j=1}^{d} \sqrt{j} \left( \text{Re}(\hat{f}(j))X_j - \text{Im}(\hat{f}(j))Y_j \right)
\]

and, hence, \( \varphi(Z_\Sigma) \) is a centered real normal random variable with variance \( \sigma_n^2 \) given by

\[
\sigma_n^2 = 4 \sum_{j=1}^{d} \frac{j}{2} \left( \text{Re}(\hat{f}(j))^2 + \text{Im}(\hat{f}(j))^2 \right) = 2 \sum_{j=1}^{d} j|\hat{f}(j)|^2.
\]

So, we obtain that

\[
\left| E[g(\varphi(Z_\Sigma))] - E[g(\sigma\zeta)] \right| = \left| E[g(\sigma_n\zeta)] - E[g(\sigma\zeta)] \right| \leq \|g\|_{Lip} \left| E[\sigma_n\zeta - \sigma\zeta] \right| \leq |\sigma - \sigma_n| \left| E[\zeta] \right| \leq |\sigma - \sigma_n|.
\]

(3.51)
3.2. Spectral properties of large random matrices from the classical compact groups

Now observe that by Proposition 3.2.27

$$|\sigma^2 - \sigma_n^2| = 2 \sum_{j=d+1}^{\infty} j |\hat{f}(j)|^2 \leq 2\|f^{(k)}\|_1^2 \sum_{j=d+1}^{\infty} \frac{1}{j^{2k-1}}$$

$$\leq 2\|f^{(k)}\|_1^2 \int_d^{\infty} \frac{1}{x^{2k-1}} dx = \frac{2\|f^{(k)}\|_1^2}{(2k-2)d^{2k-2}} = \frac{\|f^{(k)}\|_1^2}{(k-1)d^{2k-2}}.$$  

Since $|\sigma^2 - \sigma_n^2| = |(\sigma - \sigma_n)(\sigma + \sigma_n)|$, we conclude that

$$|\sigma - \sigma_n| = \frac{|\sigma^2 - \sigma_n^2|}{\sigma + \sigma_n} \leq \frac{|\sigma^2 - \sigma_n^2|}{\sigma}$$

and the third term in (3.48) can thus be bounded as follows:

$$\left| E[g(\varphi(Z_\Sigma))] - E[g(\sigma \zeta)] \right| \leq \frac{\|f^{(k)}\|_1^2}{\sigma(k-1)d^{2k-2}} \tag{3.52}$$

By (3.48), (3.49), (3.50) and (3.52) we can conclude that

$$\left| E[g(S)] - E[g(\sigma \zeta)] \right| \leq C \max \left\{ \frac{1}{d^{k-1}}, \frac{d^{7/2}}{n} \right\} \leq C \left( \frac{1}{d^{k-1}} + \frac{d^{7/2}}{n} \right), \tag{3.53}$$

where the constant $C$ only depends on $f$ and $k$. By elementary calculus one may verify, that the function $h : (0, \infty) \to \mathbb{R}$ with

$$h(x) := \frac{1}{x^{k-1}} + \frac{x^{7/2}}{n}$$

assumes its global minimum at the point

$$x_0 := \left( \frac{2(k-1)}{7} \right)^{2/7} n^{2/7},$$

and that $h$ is strictly decreasing on $(0, x_0)$ and is strictly increasing on $(x_0, \infty)$. Now, let $d := \lfloor x_0 \rfloor$. Certainly, there are $0 < \alpha < \beta$, depending only on $k$, such that for all $n \in \mathbb{N}$

$$\alpha n^{2/7} \leq d \leq \beta n^{2/7}.$$
Hence, from (3.53) we see that

\[
\left| E[g(S)] - E[g(\sigma \zeta)] \right| \leq C \left( \frac{1}{\alpha^{k-1} n^{2k-2}} + \frac{1}{n^{2k-2} n^{\frac{7}{2} (7n-1)}} \right)
\]

\[
= C \left( \frac{1}{\alpha^{k-1} n^{2k-2}} + \frac{1}{n^{2k-2} n^{\frac{2k-2}{2k+1}}} \right)
\]

\[
= \gamma n^{-\frac{2k-2}{2k+1}},
\]

where \( \gamma := \frac{1}{\alpha^{k-1}} + \beta^{7/2} \). This concludes the proof of Theorem 3.2.28.
A. Appendix

In this appendix chapter we state and prove a general version of one of de l'Hôpital's rules, which can be applied to absolutely continuous functions. Furthermore, we give a short account of the so-called Gibbs sampling procedure, which is a common tool within Stein's method for obtaining exchangeable pairs, if the random variable considered is a sum of other random variables.
A. Appendix

A.1. A general version of de l’Hôpital’s rule

The purpose of this appendix section is to prove a version of de l’Hôpital’s rule, which is general enough for all our applications of this well-known theorem. In short, the condition of differentiability will be replaced by absolute continuity, which is more natural in many measure theoretic contexts.

Lemma A.1.1. Let $a < b$ be real numbers and let $f : [a, b] \to \mathbb{R}$ be a function having the following properties:

(a) $f$ is continuous on $[a, b]$.

(b) $f|_{[c,b]}$ is absolutely continuous for each $a < c < b$ (and, hence, $f$ is $\lambda$-almost everywhere differentiable on $(a, b]$).

(c) The derivative $f'$ as a function on $(a, b)$ is essentially bounded on $(a, a + \delta)$ for some $\delta > 0$.

Then $f$ is absolutely continuous on $[a, b]$.

Proof. By the fundamental theorem of calculus for the Lebesgue integral (see for example [HS75]) it suffices to show, that there is an integrable function $g : [a, b] \to \mathbb{R}$ such that $f(x) = f(a) + \int_a^x g(t)dt$ for all $x \in [a, b]$. We put

$$g(x) = \begin{cases} f'(x) & , x \in (a, b) \\ 0 & , x = a \end{cases}$$

(A.1)

Then $g$ is a measurable function and by (c) there is an $M \in (0, \infty)$ such that $|g(t)| \leq M$ for $\lambda$-almost all $t \in (a, a + \delta)$. Thus

$$\int_a^b |g(t)|dt = \int_{[a,a+\delta]} |g(t)|dt + \int_{[a+\delta,b]} |g(t)|dt \leq \delta M + \int_{[a+\delta,b]} |f'(t)|dt < \infty,$$

since $f'$ is integrable on $[a + \delta, b]$ by (b). Thus, $g \in \mathcal{L}^1([a, b])$ and by (b) for each $x \in (a, b)$ and sufficiently large $n \in \mathbb{N}$

$$f(x) = f\left(a + \frac{1}{n}\right) + \int_{\left[a + \frac{1}{n}, x\right]} g(t)dt.$$  \hspace{1cm} (A.2)

Here the right hand side converges to $f(a) + \int_a^x g(t)dt$, since $f$ is continuous at $a$ by (a) and by the dominated convergence theorem, which applies because $g$ is integrable over $[a, b]$. This proves the lemma, since for $x = a$ the desired identity holds trivially.  \hfill \Box
One can easily see that none of the conditions (a)-(c) in Lemma A.1.1 can be dropped without substitution: Conditions (a) and (b) are quite natural and condition (c) ensures that \( f \) is of bounded variation on \([a, b]\) which is well-known to be necessary for absolute continuity.

**Lemma A.1.2.** Let \( a < b \) be real numbers and let \( f : [a, b] \to \mathbb{R} \) be a function having the following properties:

(a) \( f \) is continuous on \([a, b]\).

(b) \( f|_{[c, b]} \) is absolutely continuous for each \( a < c < b \) (and, hence, \( f \) is \( \lambda \)-almost everywhere differentiable on \((a, b]\)).

(c) There is some \( a < d < b \), a set \( A \subseteq (a, d) \) at each of whose points \( f \) is differentiable with \( \lambda ((a, d) \setminus A) = 0 \) and a real number \( \gamma \) such that \( \lim_{n \to \infty} f'(x_n) = \gamma \) for each sequence \((x_n)_{n \in \mathbb{N}}\) lying in \( A \) with \( \lim_{n \to \infty} x_n = a \).

Then \( f \) is absolutely continuous on \([a, b]\) and is differentiable at \( a \) with \( f'(a) = \gamma \).

**Proof.** By (c) the function \( f' \) is essentially bounded in a neighbourhood of \( a \). Thus, by Lemma A.1.1 \( f \) is absolutely continuous on all of \([a, b]\). To prove the last assertion, note that for \( x \in (a, d) \)

\[
\left| \frac{f(x) - f(a)}{x - a} - \gamma \right| = \left| \frac{1}{x - a} \int_a^x f'(t) dt - \gamma \right| = \left| \frac{1}{x - a} \int_a^x (f'(t) - \gamma) dt \right| \\
\leq \frac{1}{x - a} \int_a^x |f'(t) - \gamma| dt \\
= \frac{1}{x - a} \int_{(a, x) \cap A} |f'(t) - \gamma| dt, \tag{A.3}
\]

where the last equation follows from (c). Given \( \varepsilon > 0 \), by (c) we can find \( \delta > 0 \) such that for all \( t \in (a, a + \delta) \cap A \) we have \( |f'(t) - \gamma| < \varepsilon \). Thus from (A.3) we see that for \( x \in (a, a + \delta) \)

\[
\left| \frac{f(x) - f(a)}{x - a} - \gamma \right| \leq \varepsilon.
\]

This proves the claim. \( \square \)

**Remark A.1.3.** (i) Note that the value of \( \gamma \) in Lemma A.1.2 does not depend on the set \( A \) in (c). This can either be seen directly or by the conclusion of the Lemma, since the value of a (right) derivative at a point is clearly unique.
A. Appendix

(ii) Lemma A.1.2 actually implies that the function \( f' \) restricted to \( A \cup \{a\} \) is continuous at \( a \).

**Lemma A.1.4.** Let \( r \in (0, \infty) \cup \{+\infty\} \) and let \( h : (0, r) \to \mathbb{R} \) be a function with the following properties:

(a) \( h \) is absolutely continuous on every compact subinterval \([c, d] \subseteq (0, r)\).

(b) \( \lim_{x \downarrow 0} h(x) = 0 \).

(c) There is some \( 0 < s < r \), a set \( A \subseteq (0, s) \) at each of whose points \( h \) is differentiable and with \( \lambda((0, s) \setminus A) = 0 \) and a real number \( \gamma \) such that \( \lim_{n \to \infty} h'(x_n) = \gamma \) for each sequence \((x_n)_{n \in \mathbb{N}}\) lying in \( A \) with \( \lim_{n \to \infty} x_n = 0 \).

Then it holds that \( \lim_{x \downarrow 0} \frac{h(x)}{x} = \gamma \).

**Proof.** Let \( b := \frac{r + s}{2} \) and define \( \hat{h} : [0, b] \to \mathbb{R} \) by

\[
\hat{h}(x) := \begin{cases} h(x), & x \in (0, b] \\ 0, & x = 0 \end{cases}
\]

Then, by (a) and (b) \( \hat{h} \) is continuous and we see, that it fulfills all the conditions of Lemma A.1.2. Hence, by this Lemma

\[
\gamma = \hat{h}'(0) = \lim_{x \downarrow 0} \frac{\hat{h}(x) - \hat{h}(0)}{x - 0} = \lim_{x \downarrow 0} \frac{h(x)}{x},
\]

as desired. \( \square \)

Now, we are in a position to prove a version of one of de l’Hôpital’s rules involving absolutely continuous rather than differentiable functions.

**Theorem A.1.5** (generalization of one of de l’Hôpital’s rules). Let \( a < b \) be extended real numbers and let \( f, g : (a, b) \to \mathbb{R} \) be functions with the following properties:

(i) If \( a < a' < b' < b \), then both, \( f \) and \( g \), are absolutely continuous on \([a', b']\).

(ii) We have \( g'(x) > 0 \) for \( \lambda \)-almost all \( x \in (a, b) \) and with \( E := \{x \in (a, b) : g'(x) \neq 0\} \) it holds that \( \lim_{n \to \infty} \frac{f'(x_n)}{g'(x_n)} = \gamma \in \mathbb{R} \) for each sequence \((x_n)_{n \in \mathbb{N}}\) in \( E \) converging to \( a \).
A.1. A general version of de l’Hôpital’s rule

If \( \lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0 \), then \( g(x) \neq 0 \) for all \( x \in (a, b) \) and

\[
\lim_{x \to a} \frac{f(x)}{g(x)} = \gamma.
\]

The same conclusion holds if \( g'(x) < 0 \) for almost all \( x \in (a, b) \) and an analogous result is true for \( \lim_{x \to a} \).

Proof. If \( a < x < y < b \), then by (i) and (ii)

\[
g(y) - g(x) = \int_x^y g'(t) dt > 0,
\]

so that \( g \) is strictly increasing. Let \( I := (a, b) \) and \( J := g(I) = (0, r) \), where \( r = \lim_{x \to a} g(x) \in (0, \infty) \cup \{+\infty\} \). Note that \( J \) is an interval because of the continuity of \( g \) by the intermediate value theorem. It is well-known, that the inverse function \( g^{-1} : J \to I \) is also continuous and strictly increasing. What is less known, but true, is that \( g^{-1} \) is also absolutely continuous on each compact sub-interval \([c, d]\) of \( J \). This follows from \( g'(x) > 0 \) for almost all \( x \in (a, b) \) either by the Banach-Zarecki-Theorem (Theorem (18.25) in [HS75]) through verifying Lusin’ s condition or directly from another theorem of Zanecki’s which states that \( g^{-1} \) is absolutely continuous on \([c, d]\) if and only if \( \{ x \in [g^{-1}(c), g^{-1}(d)] : g'(x) = 0 \} \) has Lebesgue-measure 0. Let \( h := f \circ g^{-1} : J \to \mathbb{R} \). Our aim is to show, that \( h \) is also absolutely continuous on each compact sub-interval \([c, d]\) \( \subseteq J \). Since a composition of absolutely continuous functions is not always again absolutely continuous, we would like to appeal to the Banach-Zarecki-Theorem: Of course, \( h \) is continuous on \( J \) and satisfies Lusin’ s condition on each compact \([c, d]\) \( \subseteq J \), since \( f \) and \( g^{-1} \) do and this property is transitive. Hence, in order to show that \( h \) is absolutely continuous on \([c, d]\) it is sufficient to prove that it is of bounded variation on \([c, d]\). This is rather easy to see since \( g^{-1} \) is strictly increasing: Let \( c = y_0 < y_1 < \ldots < y_{m-1} < y_m = d \) be a partition of \([c, d]\). If \( x_j := g^{-1}(y_j), j = 0, \ldots, m \), then by the monotonicity of \( g^{-1}, a' := x_0 < x_1 < \ldots < x_{m-1} < x_m =: b' \) is a partition of \([a', b']\) and hence

\[
\sum_{j=1}^{m} |h(y_j) - h(y_{j-1})| = \sum_{j=1}^{m} |(f \circ g^{-1})(y_j) - (f \circ g^{-1})(y_{j-1})| = \sum_{j=1}^{m} |(f(x_j) - f(x_{j-1})| \leq \text{Var} (f; [a', b']),
\]

the total variance of \( f \) on \([a', b']\). Thus, also \( \text{Var} (h; [c, d]) \leq \text{Var} (f; [a', b']) < \infty \) since \( f \) as an absolutely continuous function on \([a', b']\) is of bounded variation there. Thus, by the Banach-Zarecki-Theorem \( h \) is absolutely continuous on each
compact sub-interval $[c, d] \subseteq J$ and is particularly differentiable at $\lambda$-almost every $y \in J = (0, r)$. By the chain rule formula and the formula for the derivative of the inverse function at any fixed point, for such a $y \in J$

$$h'(y) = f'(g^{-1}(y)) \cdot (g^{-1})'(y) = \frac{f'(g^{-1}(y))}{g'(g^{-1}(y))}.$$  

Let $F := E \cap \{ x \in (a, b) : f'(x) \text{ exists} \}$ and let $A := g(F) \subseteq J$. Then $h'(y)$ exists for $y \in A$ and by Lusin’s condition for $g$ we have

$$\lambda(J \setminus A) = \lambda(J \setminus g(F)) = \lambda(g(I \setminus F)) = 0,$$

since $\lambda(I \setminus F) = 0$. Now let $(y_n)_{n \in \mathbb{N}}$ be a sequence in $A$ with $\lim_{n \to \infty} y_n = 0$ and let $x_n := g^{-1}(y_n), n \in \mathbb{N}$. Then $(x_n)_{n \in \mathbb{N}}$ is a sequence in $E$ converging to $a$ and so by (ii)

$$\lim_{n \to \infty} h'(y_n) = \lim_{n \to \infty} f'(g^{-1}(y_n)) = \lim_{n \to \infty} \frac{f'(x_n)}{g'(g^{-1}(y_n))} = \gamma.$$

Since also

$$\lim_{y \to 0} h(y) = \lim_{y \to 0} f(g^{-1}(y)) = \lim_{x \to a} f(x) = 0$$

by hypothesis, we have shown, that $h$ satisfies all the assumptions of Lemma A.1.4 and so we conclude that

$$\gamma = \lim_{y \to 0} \frac{h(y)}{y} \underset{x=g(y)}{=} \lim_{x \to a} \frac{f(g^{-1}(g(x)))}{g(x)} = \lim_{x \to a} \frac{f(x)}{g(x)},$$

which was to be shown.  

\[\square\]

**Remark A.1.6.** (a) By replacing Lemma A.1.4 with another simple lemma, one can similarly prove, that the conclusion concerning the limit value of Theorem A.1.5 also holds true, if the condition $\lim_{x \to a} f(x) = \lim_{x \to a} g(x) = 0$ is replaced by $\lim_{x \to a} g(x) = -\infty$ (and the usual variants are, of course, also true) as could be guessed from the classical theorem for everywhere differentiable functions (see for example [Rud76]). This is the second one of de l’Hôpital’s rules.

(b) The conclusion of the classical theorem in [Rud76] also holds, if $\gamma = \pm \infty$. This result can also be generalized to our situation (under both hypotheses) by proving an appropriate substitute for Lemma A.1.4.
A.1. A general version of de l'Hôpital's rule

(c) There is already some literature about more general versions of de l'Hôpital's rules (see for example [Lee77], [Lee84] and [Ost76]). Particularly, the corollary in Section 2 of [Lee84] seems to be a slightly more general variant of our Theorem A.1.5 and its ramifications through the current remark, while [Lee77] involves some complicated classes of functions and concepts of limits, which are all but not well-known.
A. Appendix

A.2. The Gibbs sampler

In this appendix we recall the so-called Gibbs sampling procedure, which gives a common technique to construct exchangeable pairs in the situation where the random variable considered is a sum of other random variables. We also present general identities for conditional moments, which naturally appear in plug-in theorems within the exchangeable pairs version of Stein’s method.

Let \((\mathcal{X}, \mathcal{F})\) be a measurable space, which we assume for technical reasons to be a polish space with \(\mathcal{F}\) the \(\sigma\)-algebra of Borel sets of \(\mathcal{X}\), and let \(X_1, \ldots, X_n\) be random variables on the same probability space with values in \(\mathcal{X}\). The Gibbs sampler is a means of finding further random variables \(X_1', \ldots, X_n'\), defined jointly with \(X_1, \ldots, X_n\) perhaps on a different probability space, such that the vectors \(X := (X_1, \ldots, X_n)\) and \(X' := (X_1', \ldots, X_n)\) are exchangeable, i.e. that \((X, X')\) has the same distribution as \((X', X)\). The procedure goes as follows:

- Let \(I\) be a random index, which is independent of \(X\) and is distributed according to an arbitrarily given distribution \(Q\) on \(\{1, \ldots, n\}\).
- Observe \(X_1 = x_1, \ldots, X_n = x_n\) and \(I = i\) and construct \(Y\) according to (a regular version of) the conditional distribution of \(X_i\) given \(X_j = x_j, j \neq i\), that is \(\mathcal{L}(Y|X_1 = x_1, \ldots, X_n = x_n, I = i) := \mathcal{L}(X_i|X_j = x_j, j \neq i)\).
- Letting \(X' := (X_1', \ldots, X_n') := (X_1, \ldots, X_{i-1}, Y, X_{i+1}, \ldots, X_n)\) we have \((X, X') \overset{D}{=} (X', X)\).

**Remark A.2.1.** It is clear that one can always construct \(X\) and \(X'\) on a joint probability space \((\Omega, \mathcal{A}, P)\) using suitable transition kernels and product spaces and one can easily check that the thus obtained pair \((X, X')\) is indeed exchangeable. The assumption that \(\mathcal{X}\) be polish guarantees that the regular conditional distributions involved actually exist. In most applications one lets \(Q\) be the uniform distribution on \(\{1, \ldots, n\}\). If, additionally, also the random variables \(X_1, \ldots, X_n\) are exchangeable, then the Gibbs sampling procedure with \(Q = \delta_1\) or \(Q = \delta_n\) is usually as efficient as with a non-trivial distribution \(Q\). Note that \(X\) and \(X'\) only differ in at most one coordinate. This property makes the Gibbs sampler an efficient tool within Stein’s method.

The following easy lemma is often useful.

**Lemma A.2.2.** Let \((X, X')\) be an exchangeable pair of random variables with values in a measurable space \((\mathcal{X}, \mathcal{F})\). Let \(f: (\mathcal{X}, \mathcal{F}) \rightarrow (\mathcal{Y}, \mathcal{G})\) be a measurable mapping into another measurable space \((\mathcal{Y}, \mathcal{G})\) and let \(W := f(X)\) and \(W' := f(X')\). Then, also the pair \((W, W')\) is exchangeable.
A.2. The Gibbs sampler

In many cases we have \( W = \sum_{i=1}^{n} X_i \) or, more generally, \( W = f(X_1, \ldots, X_n) \), where \( f \) is a symmetric function, i.e. \( f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \) for each permutation \( \sigma \) of \( \{1, \ldots, n\} \).

The following proposition summarizes formulae for certain conditional moments, which usually appear in plug-in theorems within the exchangeable pairs approach of Stein’s method.

Proposition A.2.3. Let \( X_1, \ldots, X_n \) be real-valued random variables and let \( X'_1, \ldots, X'_n \) be constructed according to the Gibbs sampler with mixing distribution \( Q \). Let \( q \) be the probability mass function corresponding to \( Q \). Define \( W := \sum_{j=1}^{n} X_j \) and \( W' := \sum_{j=1}^{n} X'_j \). Then, by Lemma A.2.2 \((W, W')\) is an exchangeable pair of real-valued random variables and the following identities hold:

(a) \( E[W' - W|X] = \sum_{i=1}^{n} q(i) (E[X_i|X_j, j \neq i] - X_i) \)

- If \( Q \) is the uniform distribution, then \( E[W' - W|X] = -\frac{1}{n}W + \frac{1}{n} \sum_{i=1}^{n} E[X_i|X_j, j \neq i] \).
- If \( X_1, \ldots, X_n \) are independent, then \( E[W' - W|X] = -\sum_{i=1}^{n} q(i)(X_i - E[X_i]) \).
- If \( X_1, \ldots, X_n \) are independent and \( Q \) is the uniform distribution, then \( E[W' - W|X] = -\frac{1}{n}W - E[W] = E[W' - W|W] \).

(b) \( E[W'|W|W] = \sum_{i=1}^{n} q(i) (E[E[X_i|X_j, j \neq i]|W] - E[X_i|W]) \)

- If \( Q \) is the uniform distribution, then \( E[W'|W|W] = -\frac{1}{n}W + \frac{1}{n} \sum_{i=1}^{n} E[E[X_i|X_j, j \neq i]|W] \).
- If \( X_1, \ldots, X_n \) are independent, then \( E[W'|W|W] = -\sum_{i=1}^{n} q(i)(E[X_i|W] - E[X_i]) \).
- If \( X_1, \ldots, X_n \) are exchangeable, then \( E[W'|W|W] = -\frac{1}{n}W + E[E[X_n|X_1, \ldots, X_{n-1}]|W] \).

(c) \( E[(W' - W)^2|X] = \sum_{i=1}^{n} q(i) (E[X_i^2|X_j, j \neq i] - 2X_i E[X_i|X_j, j \neq i] + X_i^2) \)

- If \( X_1, \ldots, X_n \) are independent, then \( E[(W' - W)^2|X] = \sum_{i=1}^{n} q(i) (E[X_i^2] - 2X_i E[X_i] + X_i^2) \).
If \( X_1, \ldots, X_n \) are exchangeable, then

\[
E[(W' - W)^2 | X] = \sum_{i=1}^{n} q(i) (h_2(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \nonumber \\
- 2X_i h_1(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) + X_i^2),
\]

where \( h_1(x_1, \ldots, x_{n-1}) = E[X_n | X_1 = x_1, \ldots, X_{n-1} = x_{n-1}] \) and \( h_2(x_1, \ldots, x_{n-1}) = E[X_n^2 | X_1 = x_1, \ldots, X_{n-1} = x_{n-1}] \).

If \( X_1, \ldots, X_n \) are exchangeable, then

\[
E[(W' - W)^2 | W] = E[E[X_n^2 | X_1, \ldots, X_{n-1}] | W] \nonumber \\
- 2E[X_n E[X_n | X_1, \ldots, X_{n-1}] | W] + E[X_n^2 | W].
\]

Proof. In order to give an idea, we will exemplarily prove two of the asserted identities, namely the general formulae for \( E[W' - W | X] \) and for \( E[(W' - W)^2 | X] \).

Note that

\[
E[W' - W | X] = E[E[W' - W | X, I] | X]. \quad \text{(A.4)}
\]

By the construction of \( X' \), for any \( x = (x_1, \ldots, x_n) \) and \( i \in \{1, \ldots, n\} \) we have

\[
E[W' - W | X = x, I = i] = E[X'_i - X_i | X = x, I = i] = E[Y | X = x, I = i] - x_i \nonumber \\
= E[X_i | X_j = x_j, j \neq i] - x_i \nonumber \\
= \sum_{k=1}^{n} 1_{(i=k)} (E[X_k | X_j = x_j, j \neq k] - x_k).
\]

This implies that

\[
E[W' - W | X, I] = \sum_{i=1}^{n} 1_{(I=i)} (E[X_i | X_j, j \neq i] - X_i).
\]

For each \( i \in \{1, \ldots, n\} \) the random variable \( E[X_i | X_j, j \neq i] - X_i \) is \( \sigma(X) \)-measurable and \( I \) is independent of \( X \). Thus, we obtain from (A.4) that
\[
E[W' - W \mid X] = \sum_{i=1}^{n} E[1_{I=i} (E[X_i \mid X_j, j \neq i] - X_i) \mid X] \\
= \sum_{i=1}^{n} (E[X_i \mid X_j, j \neq i] - X_i) E[1_{I=i} \mid X] \\
= \sum_{i=1}^{n} P(I = i) (E[X_i \mid X_j, j \neq i] - X_i) \\
= \sum_{i=1}^{n} q(i) (E[X_i \mid X_j, j \neq i] - X_i),
\]

as claimed.

Similarly to (A.4), we have

\[
E[(W' - W)^2 \mid X] = E[E[(W' - W)^2 \mid X, I] \mid X]. \tag{A.5}
\]

Again, by construction, for any \(x = (x_1, \ldots, x_n)\) and \(i \in \{1, \ldots, n\}\) we have

\[
E[(W' - W)^2 \mid X = x, I = i] = E[(X'_i - X_i)^2 \mid X = x, I = i] \\
= E[Y^2 \mid X = x, I = i] - 2x_i E[Y \mid X = x, I = i] + x_i^2 \\
= E[X_i^2 \mid X_j = x_j, j \neq i] - 2x_i E[X_i \mid X_j = x_j, j \neq i] + x_i^2 \\
= \sum_{k=1}^{n} 1_{i=k} \left( E[X_k^2 \mid X_j = x_j, j \neq k] - 2x_k E[X_k \mid X_j = x_j, j \neq k] + x_k^2 \right).
\]

This implies that

\[
E[(W' - W)^2 \mid X, I] \\
= \sum_{i=1}^{n} 1_{I=i} \left( E[X_i^2 \mid X_j, j \neq i] - 2x_i E[X_i \mid X_j, j \neq i] + x_i^2 \right).
\]

Arguing as before, from (A.5) we obtain
A. Appendix

\[
E[(W' - W)^2 \mid X] \\
= \sum_{i=1}^{n} \left( E[X_i^2 \mid X, j \neq i] - 2X_i E[X_i \mid X, j \neq i] + X_i^2 \right) E[1_{i=\hat{i}} \mid X] \\
= \sum_{i=1}^{n} q(\hat{i}) \left( E[X_i^2 \mid X, j \neq \hat{i}] - 2X_i E[X_i \mid X, j \neq \hat{i}] + X_i^2 \right).
\]

\[\square\]
Bibliography


Bibliography


Bibliography


Bibliography


Bibliography


