Lecture Notes in

STEIN'S METHOD

Martin Raič

Department of Statistics and Applied Probability
National University of Singapore

Last change: April 18, 2020

Contents

Chapter 1

Introduction

1.1 Motivation and outline

Stein's method serves as a powerful tool for estimating the error in approximating complicated probability distributions by more tractable ones. It was introduced in 1970 by Charles Stein (1920–2016) [31], first for assessing the error in the central limit theorem. Recall that, roughly speaking, if $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with mean μ_1 and variance σ_1^2, then:

$$X_1 + X_2 + \cdots + X_n \approx \mathcal{N}(n\mu_1, n\sigma_1^2)$$

or equivalently,

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu_1}{\sigma_1\sqrt{n}} \approx \mathcal{N}(0,1)\,.$$

Keeping μ_1 and σ_1^2 fixed, it is known that the larger n, the smaller is the error in the preceding approximation. Although the error was estimated well before Stein, most prominently by Berry [7] and Esseen [16], Stein's method allows for numerous generalizations where classical approaches, such as characteristic functions, seem not to work. In particular, Stein's method succeeds to cope successfully with many sorts of dependence.

Stein's method has been adapted to numerous other approximations. The first such modification seems to be due to Louis H Y Chen [9], who adapted it to Poisson approximation. Recall that if $\lambda > 0$ and $X_1, X_2, \ldots$ are independent and identically distributed random variables following the Bernoulli distribution $\mathrm{Be}(\lambda/n) = \begin{pmatrix} 0 & 1 \\ 1-\lambda/n & \lambda/n \end{pmatrix}$, then:

$$X_1 + X_2 + \cdots + X_n \approx \mathrm{Po}(\lambda)\,.$$

There are several surveys of Stein's method. Stein summarizes his work in his book [32]. For more modern surveys, see Barbour [2], and Barbour and Chen [3, 4].

The idea behind Stein's method could be viewed as follows: let $\mathscr{M}$ be a vector space of certain signed measures containing a tractable probability measure ν. To approximate

other (presumably complicated) measures $\mu \in \mathscr{M}$ by ν, we first find a linear operator $\mathcal{B}\colon \mathscr{M} \to \mathscr{Y}$ which annihilates ν, i. e.,

$$\mathcal{B}\nu = 0\,;$$

here, $\mathscr{Y}$ is another vector space. We strive to measure the size of the error $\mu - \nu$ in terms of the size of $\mathcal{B}\mu$.

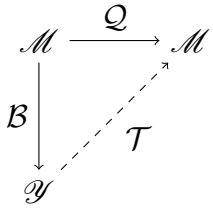
The equality $\mathcal{B}\nu = 0$ can be rewritten as $\nu \in \ker \mathcal{B}$. If $\mu - \nu$ can be estimated in terms of $\mathcal{B}\mu$, then $\mathcal{B}$ should separate probability measures sufficiently well, so that it is reasonable to assume that $\ker \mathcal{B} = \operatorname{span}(\{\nu\})$. Denoting

$$\langle h\,,\,\rho\rangle := \int h \,\mathrm{d}\rho \tag{1.1.1}$$

(provided that $h \in L^1(|\rho|)$), observe that the operator

$$\mathcal{Q}\rho := \rho - \langle 1\,,\,\rho\rangle\,\nu$$

has the very same kernel. Therefore, looking algebraically, there exists a linear operator $\mathcal{T}\colon \mathscr{Y} \to \mathscr{M}$, such that $\mathcal{T}\mathcal{B} = \mathcal{Q}$:



Remark 1.1.1. The operator $\mathcal{Q}\rho := \rho - \langle 1\,,\,\rho\rangle\,\nu$ is not the only option, but in the present notes, we shall not consider other possibilities.

If μ is a probability measure, then $\mathcal{Q}\mu = \mathcal{T}\mathcal{B}\mu = \mu - \nu$. Therefore, in order to bound the size of $\mu - \nu$, we can first bound the size of $\mathcal{B}\mu$ and then the size of $\mathcal{T}\mathcal{B}\mu$.

By size, we here mean a suitable seminorm. Seminorms are often based on a certain class of test functions $\mathscr{H}_1$:

$$\|\rho\| = \sup_{h\in\mathscr{H}_1} \big|\langle h\,,\,\rho\rangle\big|$$

(provided that $h \in L^1(|\rho|)$ for all $h \in \mathscr{H}_1$): see Appendix A.

Suppose that $\mathcal{B}$ is dual to some operator $\mathcal{A}$. That is, suppose that:

- $\mathscr{H}_1 \subseteq \mathscr{H}$, where $\mathscr{H}$ is another vector space, such that $h \in L^1(|\rho|)$ for all $h \in \mathscr{H}$;
- there is another vector space $\mathscr{F}$ and a bilinear function from $\mathscr{F} \times \mathscr{Y}$ to $\mathbb{R}$ again denoted by $\langle\cdot\,,\,\cdot\rangle$;
- $\mathcal{A}\colon \mathscr{F} \to \mathscr{H}$ is such that $\langle f\,,\,\mathcal{B}\rho\rangle = \langle \mathcal{A}f\,,\,\rho\rangle$ for all $f \in \mathscr{F}$ and $\rho \in \mathscr{M}$.

Remark 1.1.2. In practice, $\mathscr{F}$ is a space of functions, though it could in general be any vector space.

Then, of course,

$$\langle \mathcal{A}f\,,\,\nu\rangle = \langle f\,,\,\mathcal{B}\nu\rangle = 0\,.$$

Seeking $\mathcal{T}$ as being dual to some operator $\mathcal{S}\colon\ \mathscr{H} \to \mathscr{F}$, observe that

$$\langle h\,,\,\mu-\nu\rangle = \langle h\,,\,\mathcal{T}\mathcal{B}\mu\rangle = \langle \mathcal{A}\mathcal{S}h\,,\,\mu\rangle$$

and, on the other hand,

$$\langle h\,,\,\mu-\nu\rangle = \big\langle h - \langle h\,,\,\nu\rangle 1\,,\,\mu\big\rangle\,.$$

Putting all together, we obtain

Outline of Stein's method. To estimate $\langle h\,,\,\mu-\nu\rangle$ for all $h\in\mathscr{H}$:

1. Find a vector space $\mathscr{F}$ and a linear operator $\mathcal{A}\colon\ \mathscr{F}\to\mathscr{H}$, such that $\langle \mathcal{A}f\,,\,\nu\rangle = 0$ for all $f\in\mathscr{F}$. The operator $\mathcal{A}$ is called *Stein operator*.
2. For each $h\in\mathscr{H}$, find $f\in\mathscr{F}$ solving the *Stein equation*

   $$\mathcal{A}f = h - \langle h\,,\,\nu\rangle 1\,.$$

   Then we have

   $$\langle h\,,\,\mu-\nu\rangle = \langle \mathcal{A}f\,,\,\mu\rangle\,.$$
3. Efficiently bound the right hand side of the preceding equation, which is called *Stein expectation*.

Remark 1.1.3. One we have found $\mathcal{A}$, we can drop the assumption that it is dual to some operator $\mathcal{B}$. This allows us to extend the space $\mathscr{M}$, i. e., the family of measures which can be approximated.

1.2 Basic examples of Stein operators

There is no unique way of finding a Stein operator. Indeed, there might be different Stein operator for the same approximating distribution ν. Below, we give two basic methods:

1. We derive the operator from the *approximating distribution* ν (provided that we already have one in mind). Typically we first derive the operator $\mathcal{B}$ by comparing consequent point probabilities (for discrete distributions) or by differentiating the density (for continuous distributions). Then observe that $\mathcal{B}$ is dual to some $\mathcal{A}$.
2. We derive the operator from the probability distribution μ, which is *to be approximated*. Let W be a random variable with distribution μ. Typically, we perturb W to W', which is defined on the same probability space and follows the same distribution μ. Often, W and W' are exchangeable, but in general, this is not necessary. Next, if $\mathcal{G}$ is a linear operator, such that

   $$\mathbb{E}\big[f(W') - f(W)\;\big|\;W\big] = \mathcal{G}f(W)\,,$$

then $\mathbb{E}\big[\mathcal{G}f(W)\big] = 0$. If $\mathcal{G}$ is tractable, we can seek $\mathcal{A} \approx \mathcal{G}$. We often rescale and seek $\mathcal{A} \approx \kappa\mathcal{G}$, where κ is a constant factor (if we consider a sequence of random variables W_n, the order of the underlying operators $\mathcal{G}_n$ often depends on n, whereas we wish the operators $\mathcal{A}_n$ to converge to a certain operator). If $\mathcal{G}$ is not tractable, we can try

$$\mathbb{E}\big[\kappa\big(f(W') - f(W)\big) - R_f \;\big|\; W\big] = \mathcal{A}f(W)\,,$$

where R_f is intended to be of smaller order than $\kappa\big(f(W') - f(W)\big)$, as we seek $\mathcal{A}$ close to $\kappa\mathcal{G}$.

It is also possible to consider several perturbations W'_α. If they are indexed over a countable set, consider factors κ_α and let

$$\mathbb{E}\bigg[\sum_\alpha \kappa_\alpha\big(f(W') - f(W)\big) - R_f \;\bigg|\; W\bigg] = \mathcal{A}f(W)\,. \tag{1.2.1}$$

The perturbations can even be indexed over a measurable space. In this case, take κ to be a measure on this space and let

$$\mathbb{E}\bigg[\int \big(f(W'_\alpha) - f(W)\big)\,\kappa(\mathrm{d}\alpha) - R_f \;\bigg|\; W\bigg] = \mathcal{A}f(W)\,. \tag{1.2.2}$$

The approximating distribution ν should be an annihilator of the image of $\mathcal{A}$, i. e., $\langle \mathcal{A}f\,,\,\nu\rangle = 0$ for all $f \in \mathscr{F}$ (by this method, $\mathscr{F}$ must be a space of functions). This method can also help solve the Stein equation. Details will be given later: see the beginning of Chapter 2.

Below we give three examples of Stein operators obtained by the first method. Applications of the second method will be given later.

1. *Poisson approximation.* Let $\lambda > 0$ and let $\nu = \mathrm{Po}(\lambda)$ be the Poisson distribution. This is a probability measure on $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$. Signed measures on $\mathbb{N}_0$ can be identified with their mass functions, so that we simply write

$$\nu(k) = \frac{\lambda^k e^{-\lambda}}{k!}\,.$$

Accordingly, for each signed measure ρ on $\mathbb{N}_0$ and each function $h \in L^1(|\rho|)$, we have

$$\langle h\,,\,\rho\rangle = \sum_{k=0}^{\infty} h(k)\,\rho(k)\,.$$

Now observe that

$$\frac{\nu(k-1)}{\nu(k)} = \frac{k}{\lambda}$$

for all $k \in \mathbb{N} := \{1, 2, 3, \ldots\}$. This can be formulated in terms of linear operators: let $\mathcal{B}$ be the operator mapping from the space $\mathscr{M}$ of signed measures ρ on $\mathbb{N}_0$ with $\sum_{k=0}^{\infty} k\,|\rho(k)| < \infty$ to the space $\mathscr{Y}$ of all signed measures on $\mathbb{N}$ defined by

$$\mathcal{B}\rho(k) := \lambda\,\rho(k-1) - k\,\rho(k)\,.$$

Then $\mathcal{B}\nu = 0$.

Now let $\mathscr{F}$ be the space of all bounded functions on $\mathbb{N}$. Observe that for each $f \in \mathscr{F}$,

$$\begin{aligned}\langle f\,,\,\mathcal{B}\rho\rangle &= \sum_{k=1}^{\infty} \lambda\, f(k)\,\rho(k-1) - \sum_{k=1}^{\infty} k\, f(k)\,\rho(k)\\ &= \sum_{w=0}^{\infty} \lambda\, f(w+1)\,\rho(w) - \sum_{w=1}^{\infty} w\, f(w)\,\rho(w)\end{aligned}$$

(notice that all sums converge due to the definitions of $\mathscr{F}$ and $\mathscr{Y}$). Let $\mathscr{H}$ be the space of all functions on $\mathbb{N}_0$, such that there exists M, such that $|h(w)| \leq M\,w$ for all $w \in \mathbb{N}_0$. Then $\mathcal{B}$ is dual to the operator $\mathcal{A}\colon\ \mathscr{F} \to \mathscr{H}$ defined by

$$\mathcal{A}f(w) := \begin{cases} \lambda\, f(w+1) - w\, f(w) & ;\ w = 1, 2, 3, \ldots \\ \lambda\, f(1) & ;\ w = 0\,. \end{cases} \tag{1.2.3}$$

This is the usual Stein operator for the Poisson distribution. To simplify computations, we shall usually assume that f is also defined at 0 in an arbitrary way.

2. *Binomial approximation.* Let $n \in \mathbb{N}$, let $0 < \theta < 1$ and let $\nu = \mathrm{Bin}(n, \theta)$ be the binomial distribution. This is a probability measure on $\{0, 1, 2, \ldots, n\}$ given by

$$\nu(k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}\,.$$

Similarly as for the Poisson distribution, observe that

$$\frac{\nu(k-1)}{\nu(k)} = \frac{k}{n-k+1}\,\frac{1-\theta}{\theta}$$

for all $k \in \{1, 2, \ldots, n\}$. Thus, if $\mathcal{B}$ be the operator mapping from the space $\mathscr{M}$ of signed measures (or, equivalently, functions) on $\{0, 1, \ldots, n\}$ to the space $\mathscr{Y}$ of all signed measures on $\{1, 2, \ldots, n\}$ defined by

$$\mathcal{B}\rho(k) := (n-k+1)\theta\,\rho(k-1) - k(1-\theta)\,\rho(k)\,,$$

we have $\mathcal{B}\nu = 0$.

Leting $\mathscr{F}$ be the space of all functions on $\{1, 2, \ldots, n\}$, observe that for each $f \in \mathscr{F}$,

$$\begin{aligned}\langle f\,,\,\mathcal{B}\rho\rangle &= \sum_{k=1}^{n} (n-k+1)\theta\, f(k)\,\rho(k-1) - \sum_{k=1}^{n} k(1-\theta)\, f(k)\,\rho(k)\\ &= \sum_{w=0}^{n-1} (n-w)\theta\, f(w+1)\,\rho(w) - \sum_{w=1}^{n} w(1-\theta)\, f(w)\,\rho(w)\,.\end{aligned}$$

Letting $\mathscr{H}$ be the space of all functions on $\{0, 1, 2, \ldots, n\}$, $\mathcal{B}$ is dual to the operator $\mathcal{A}\colon\ \mathscr{F} \to \mathscr{H}$ defined by

$$\mathcal{A}f(w) := \begin{cases} (n-w)\theta\, f(w+1) - w(1-\theta)\, f(w) & ;\ w = 1, 2, 3, \ldots, n-1 \\ (n-w)\theta\, f(1) & ;\ w = 0 \\ -n(1-\theta)\, f(w) & ;\ w = n\,. \end{cases}$$

This is one of possible choices for the Stein operator for the binomial distribution. Again, to simplify computations, we can assume that f is also defined at 0 and $n+1$ in an arbitrary way.

3. *Normal and other continuous approximations.* Let $\nu = \mathcal{N}(0, 1)$ be the standard normal distribution (the generalization to arbitrary mean and variance is straightforward). This is a continuous distribution, and we shall identify continuous distributions with their densities (more precisely, with equivalence classes of functions, where two functions are equivalent if they differ on a set with measure zero). Thus, we have

$$\nu(x) = e^{-x^2/2}\,,$$

which is differentiable with

$$\nu'(x) = -x\, e^{-x^2/2}\,.$$

Thus, if, let's say, $\mathscr{M}$ is the space of all continuously differentiable functions on the real line with exponentially decaying derivative, $\mathscr{Y}$ is the space of all continuous exponentially decaying functions, and $\mathcal{B}\colon\ \mathscr{M} \to \mathscr{Y}$ is defined by

$$\mathcal{B}\rho(x) := \rho'(x) + x\,\rho(x)\,,$$

then $\mathcal{B}\nu = 0$. Let $\mathscr{F}$ be the space of all continuously differentiable functions on the real line of polynomial growth. For each $f \in \mathscr{F}$ and $\rho \in \mathscr{M}$, integration by parts gives

$$\begin{aligned} \langle f\,,\, \mathcal{B}\rho\rangle &= \int_{-\infty}^{\infty} f(x)\,\rho'(x)\,\mathrm{d}x + \int_{-\infty}^{\infty} x\, f(x)\,\rho(x)\,\mathrm{d}x \\ &= -\int_{-\infty}^{\infty} f'(x)\,\rho(x)\,\mathrm{d}x + \int_{-\infty}^{\infty} x\, f(x)\,\rho(x)\,\mathrm{d}x \end{aligned}$$

(polynomial growth of f and exponential decay of ρ ensure that $\lim_{x\to\pm\infty} f(x)\,\rho(x) = 0$). Letting $\mathscr{H}$ be the space of all functions on the real line of polynomial growth, we find that $\mathcal{B}$ is dual to the operator $\mathcal{A}\colon\ \mathscr{F} \to \mathscr{H}$ defined by

$$\mathcal{A}f(w) := -f'(x) + x\, f(x)\,.$$

However, usually, we take $\mathcal{A}$ with the opposite sign, i. e., $\mathcal{A}f(w) = f'(x) - x\, f(x)$.

Remark 1.2.1. In view of Remark 1.1.3, the preceding derivation of the Stein operator for the normal distribution does not imply that normal approximation is restricted to measures with continuously differentiable densities. Indeed, the space $\mathscr{M}$ can be extended to quite a general class of measures, but the extension can depend on the class $\mathscr{H}$ of test functions.

Remark 1.2.2. Under suitable conditions, for a probability density ν with continuously differentiable derivative and with $\nu'(x) = \psi(x)\,\nu(x)$, we obtain a Stein operator $\mathcal{A}f(x) = f'(x) - \psi(x)\,f(x)$. We omit the details.

Chapter 2

Poisson approximation

2.1 Independent trials

Let $X_i \sim \mathrm{Be}(p_i)$, $i \in \mathscr{I}$, be independent Bernoulli random variables, i. e., $\mathbb{P}(X_i = 1) = p_i$ and $\mathbb{P}(X_i = 0) = 1 - p_i$ for all $i \in \mathscr{I}$. Assume that $\sum_{i\in\mathscr{I}} p_i = \lambda < \infty$. Notice that due to the latter condition, we can assume without loss of generality that $\mathscr{I}$ is countable (but may be infinite). By a well known result, the sum $W = \sum_{i\in\mathscr{I}} X_i$ converges almost surely.

It is known that if, roughly speaking, the p_i's are small, then W approximately follows the Poisson distribution. However, we shall here pretend that we do not know this fact and try to find a Stein operator by the second method mentioned at the beginning of Section 1.2. Actually, at some point, we use the fact that if $\mathscr{I} = \{1, 2, \dots, n\}$, $p_i = \lambda/n$ for all i and n tends to the infinity, then W weakly converges to some distribution, but we will not need to know to which one.

Let the collection $(X_i')_{i\in\mathscr{I}}$ be an independent copy of the collection $(X_i)_{i\in\mathscr{I}}$. Denote $W_i := W - X_i = \sum_{j\in\mathscr{I}\setminus\{i\}} X_j$. Then $W_i' := W_i + X_i'$ follows the same distribution as W. In fact, W and W' even form an exchangeable pair, but we shall not need this fact.

Define a relation $U \sim V$ iff $\mathbb{E}(U \mid W) = \mathbb{E}(V \mid W)$. In addition, let $\Delta f(w) := f(w+1) - f(w)$ be the forward difference. Now take a function $g\colon\ \mathbb{N}_0 \to \mathbb{R}$ and observe that

$$\begin{aligned} g(W_i') - g(W) &= \big(g(W_i') - g(W_i)\big) - \big(g(W) - g(W_i)\big) \\ &= \Delta g(W_i)X_i' - \Delta g(W-1)X_i \\ &\sim p_i\,\Delta g(W_i) - \Delta g(W-1)X_i \\ &= p_i\,\Delta g(W) - p_i\,\Delta^2 g(W-1)X_i - \Delta g(W-1)X_i \end{aligned}$$

with the following two supplements:

- The calculation is valid provided that $\Delta^2 g$ is bounded: if $|\Delta^2 g(w)| \le M$ for all $w \in \mathbb{N}_0$, then $|\Delta g(W_i) - \Delta g(W)| \le |\Delta^2 g(W)| \le M$. Thus, $\mathbb{E}\big(|\Delta g(W_i)| \;\big|\; W\big) \le |\Delta g(W)| + M$ and $\mathbb{E}\big(|\Delta g(W_i)|X_i \;\big|\; W\big) \le p_i\big(|\Delta g(W)| + M\big)$.
- For the sake of simplicity, we here assume that g is also defined at -1; the values of the expressions are independent of the choice of $g(-1)$.

Now, in view of (1.2.1), sum over i:

$$\sum_{i\in\mathscr{I}}\big(g(W_i')-g(W)\big) \sim \lambda\,\Delta g(W) - \sum_{i\in\mathscr{I}} p_i\,\Delta^2 g(W-1)X_i - \Delta g(W-1)W$$
$$= \tilde{\mathcal{A}}g(W) - \sum_{i\in\mathscr{I}} p_i\,\Delta^2 g(W-1)X_i\,,$$

where $\tilde{\mathcal{A}}g(w) := \lambda\,\Delta g(w) - \Delta g(w-1)\,w$, noting that if $\Delta^2 g$ is bounded, the summation is valid as well (in particular, all expressions almost surely exist). As a result, we have

$$\mathbb{E}\big[\tilde{\mathcal{A}}g(W)\big] = \sum_{i\in\mathscr{I}} p_i\,\mathbb{E}\big[\Delta^2 g(W-1)X_i\big]\,,$$

From Section 1.2, recall that it makes sense to choose $\tilde{\mathcal{A}}$ so that $\sum_{i\in\mathscr{I}} p_i\,\Delta^2 g(W-1)X_i$ is of smaller order than $\tilde{\mathcal{A}}g(W)$. In our case, if $|\Delta^2 g(w)| \le M$ for all $w\in\mathbb{N}_0$ (but not necessarily $w=-1$), we can estimate:

$$\sum_{i\in\mathscr{I}} p_i\,\mathbb{E}\big|\Delta^2 g(W-1)X_i\big| \le M\sum_{i\in\mathscr{I}} p_i^2\,.$$

Now if $\mathscr{I}=\{1,2,\dots,n\}$, $p_i=\lambda/n$ for all i and n tends to the infinity, then recall that W weakly converges to some distribution, or, say, random variable. Then it is plausible that $\tilde{\mathcal{A}}g(W)$ also converges in distribution to some random variable, while $\sum_{i\in\mathscr{I}} p_i^2 = \lambda^2/n$ tends to zero. Thus, correct terms have been stored into $\tilde{\mathcal{A}}g(W)$.

We should now seek a probability measure ν which annihilates the image of $\tilde{\mathscr{A}}$. However, we can just observe that $\tilde{\mathcal{A}}g = \mathcal{A}f$, where $f(w)=\Delta g(w-1)$ and where $\mathcal{A}$ is defined as in (1.2.3), that is

$$\mathcal{A}f(w) = \lambda\,f(w+1) - w\,f(w)$$

(again, we assume that $f(0)$ is defined, but for $w\in\mathbb{N}_0$, the value of $\mathcal{A}f(w)$ is independent of $f(0)$). In Section 1.2, we have shown that the Poisson distribution $\mathrm{Po}(\lambda)$, that is,

$$\mathrm{Po}(\lambda)(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

annihilates $\mathcal{A}f$ (that is, $\big\langle \mathcal{A}f\,,\,\mathrm{Po}(\lambda)\big\rangle = 0$), provided that f is bounded. However, it is straightforward to check that this remains true if Δf is bounded. Therefore, $\big\langle \tilde{\mathcal{A}}g\,,\,\mathrm{Po}(\lambda)\big\rangle = 0$ if $\Delta^2 g$ is bounded.

For each $f\colon\,\mathbb{N}_0\to\mathbb{R}$, such that $|\Delta f(w)|\le M$ for all $w\in\mathbb{N}$ (but not necessarily $w=0$), there exists $g\colon\,\{-1,0,1,\dots\}\to\mathbb{R}$, such that $f(w)=\Delta g(w-1)$ for all $w\in\mathbb{N}_0$. Therefore, for each f satisfying the above-mentioned condition, we have

$$\mathbb{E}\big[\mathcal{A}f(W)\big] = \sum_{i\in\mathscr{I}} p_i\Delta f(W)X_i$$

(and it is also straightforward to verify this identity directly). If f is a solution to the Stein equation

$$\mathcal{A}f(w) = h(w) - \big\langle h\,,\,\mathrm{Po}(\lambda)\big\rangle\,, \tag{2.1.1}$$

this gives a bound on the error in the Poisson approximation, as summarized in the assertion below.

Proposition 2.1.1. *If f is a solution to (2.1.1) with $|\Delta f(w)| \leq M$ for all $w \in \mathbb{N}$ (but not necessarily $w = 0$), then we have*

$$\left|\mathbb{E}\big[h(W)\big] - \big\langle h\,,\,\mathrm{Po}(\lambda)\big\rangle\right| \leq M \sum_{i \in \mathscr{I}} p_i^2\,. \tag{2.1.2}$$

□

2.2 Solution to the Stein equation

In the previous section, we have proved that we can efficiently estimate the error in the Poisson approximation of the sum of independent Bernoulli indicators with small success probabilities provided that we can bound Δf for some solution f to the Stein equation (2.1.1). This is what we shall do in this section. However, the bounds obtained here will not only be useful for sums of independent indicators, but also for indicators with certain dependence structure, which will be described in the sequel.

We shall here bound Δf for the case where h is an indicator of a set $A \subseteq \mathbb{N}_0$. This will allow us to bound the difference $\mathbb{P}(W \in A) - \mathrm{Po}(\lambda)(A)$.

First, we shall find a solution f_j for the indicator of the singleton $\{j\}$, i. e., the solution to the equation

$$\lambda\, f_j(w+1) - w\, f_j(w) = \mathbf{1}(w = j) - \frac{\lambda^j}{j!}\, e^{-\lambda}\,. \tag{2.2.1}$$

Seeking f_j to be in the form

$$f_j(w) = \frac{(w-1)!}{\lambda^w}\, \psi_j(w) \qquad \text{for all } w \in \mathbb{N},$$

equation (2.2.1) reduces to

$$\psi_j(w+1) - \psi_j(w) = \frac{\lambda^w}{w!}\left(\mathbf{1}(w = j) - \frac{\lambda^j}{j!}\, e^{-\lambda}\right) \qquad \text{for all } w \in \mathbb{N}_0, \tag{2.2.2}$$

setting $\psi_j(0) := 0$. For $w = 1, 2, \ldots, j$, we obtain

$$\psi_j(w) = \sum_{k=0}^{w-1} \big(\psi_j(k+1) - \psi_j(k)\big) = -\frac{\lambda^j}{j!}\, e^{-\lambda} \sum_{k=0}^{w-1} \frac{\lambda^k}{k!}\,. \tag{2.2.3}$$

Next, if we wish Δf_j to be bounded, we must have $\lim_{w\to\infty} \psi_j(w) = 0$. But then we have

$$\psi_j(w) = -\sum_{k=w}^{\infty} \big(\psi_j(k+1) - \psi_j(k)\big) = \frac{\lambda^j}{j!}\, e^{-\lambda} \sum_{k=w}^{\infty} \frac{\lambda^k}{k!} \tag{2.2.4}$$

for all $w = j+1, j+2, \ldots$ The function ψ_j is now completely determined and satisfies equation (2.2.2) for all $w \notin j$. For $w = j$, we obtain

$$\begin{aligned}
\psi_j(j+1) - \psi_j(j) &= \frac{\lambda^j}{j!}\, e^{-\lambda} \left(\sum_{k=0}^{j-1} \frac{\lambda^k}{k!} + \sum_{k=j+1}^{\infty} \frac{\lambda^k}{k!} \right) \\
&= \frac{\lambda^j}{j!}\, e^{-\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} - \frac{\lambda^j}{j!} \right) \\
&= \frac{\lambda^j}{j!}\, e^{-\lambda} \left(e^{\lambda} - \frac{\lambda^j}{j!} \right) \\
&= \frac{\lambda^j}{j!} \left(1 - \frac{\lambda^j}{j!}\, e^{-\lambda} \right)
\end{aligned}$$

and (2.2.2) holds true in this case, too.

Now we examine $\Delta f_j(w)$. We distinguish three cases.

Case 1: $w = 1, 2, \ldots, j-1$. Write

$$\begin{aligned}
f_j(w+1) &= -\frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{w!}{\lambda^{w+1}} \sum_{l=0}^{w} \frac{\lambda^l}{l!}\,, \\
f_j(w) &= -\frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{(w-1)!}{\lambda^{w}} \sum_{k=0}^{w-1} \frac{\lambda^k}{k!} \\
&= -\frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{(w-1)!}{\lambda^{w}} \sum_{l=1}^{w} \frac{\lambda^{l-1}}{(l-1)!} \\
&= -\frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{(w-1)!}{\lambda^{w+1}} \sum_{l=0}^{w} \frac{l\,\lambda^l}{l!}\,, \\
\Delta f_j(w) &= -\frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{(w-1)!}{\lambda^{w+1}} \sum_{l=0}^{w} \frac{(w-l)\lambda^l}{l!} \\
&\leq 0\,.
\end{aligned}$$

Case 2: $w = j+1, j+2, \ldots$ Write

$$\begin{aligned} f_j(w+1) &= \frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{w!}{\lambda^{w+1}} \sum_{l=w+1}^{\infty} \frac{\lambda^l}{l!}\,, \\ f_j(w) &= \frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{(w-1)!}{\lambda^{w}} \sum_{k=w}^{\infty} \frac{\lambda^k}{k!} \\ &= \frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{(w-1)!}{\lambda^{w}} \sum_{l=w+1}^{\infty} \frac{\lambda^{l-1}}{(l-1)!} \\ &= \frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{(w-1)!}{\lambda^{w+1}} \sum_{l=w+1}^{\infty} \frac{l\,\lambda^l}{l!}\,, \\ \Delta f_j(w) &= \frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{(w-1)!}{\lambda^{w+1}} \sum_{l=w+1}^{\infty} \frac{(w-l)\,\lambda^l}{l!} \\ &\leq 0\,. \end{aligned}$$

Case 3: $w = j$. Write

$$\begin{aligned} f_j(j) &= -\frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{(j-1)!}{\lambda^{j}} \sum_{k=0}^{j-1} \frac{\lambda^k}{k!} = -\frac{e^{-\lambda}}{j} \sum_{l=1}^{j} \frac{\lambda^{l-1}}{(l-1)!}\,, \\ f_j(j+1) &= \frac{\lambda^j}{j!}\, e^{-\lambda}\, \frac{j!}{\lambda^{j+1}} \sum_{k=j+1}^{\infty} \frac{\lambda^k}{k!} = \frac{e^{-\lambda}}{\lambda} \sum_{l=j+1}^{\infty} \frac{\lambda^l}{l!}\,, \\ \Delta f_j(j) &= \frac{e^{-\lambda}}{\lambda} \left(\frac{1}{j} \sum_{l=1}^{j} \frac{\lambda^l}{(l-1)!} + \sum_{l=j+1}^{\infty} \frac{\lambda^l}{l!} \right) \leq \frac{e^{-\lambda}}{\lambda} \sum_{l=1}^{\infty} \frac{\lambda^l}{l!} = \frac{1-e^{-\lambda}}{\lambda}\,. \end{aligned}$$

Therefore, $\Delta f_j(w) \leq \frac{1-e^{-\lambda}}{\lambda}$ for all $w \in \mathbb{N}$.

For convenience, set $f_j(0) := 0$ for all $j \in \mathbb{N}_0$. Now take $A \subseteq \mathbb{N}_0$ and define $f_A(w) := \sum_{j\in A} f_j(w)$. We will show that the series converges pointwise and that f_A solves the Stein equation

$$\lambda\, f_A(w+1) - w\, f_A(w) = \mathbf{1}(w \in A) - \mathrm{Po}(\lambda)(A)\,. \tag{2.2.5}$$

Clearly, $\mathbf{1}(w \in A) - \mathrm{Po}(\lambda)(A) = \sum_{j\in A} \big(\mathbf{1}(w=j) - \mathrm{Po}(\lambda)(\{j\})\big)$. As to the left hand side, first deduce from (2.2.3) and (2.2.4) that $|\psi_j(w)| \leq \frac{\lambda^j}{j!}\, e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \frac{\lambda^j}{j!}$ for all $w \in \mathbb{N}_0$. Therefore,

$$\sum_{j=0}^{\infty} |f_j(w)| \leq \frac{(w-1)!}{\lambda^w} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \frac{(w-1)!}{\lambda^w}\, e^{\lambda} < \infty\,.$$

As a result, the series $\sum_{j\in A} f_j(w)$ converges pointwise and, moreover, $\lambda f_A(w+1) - w\, f_A(w) = \sum_{j\in A} \big(\lambda f_j(w+1) - w\, f_j(w)\big)$. Combining all together, the proof of (2.2.5) is complete.

We have now shown that $\Delta f_A(w) \leq \frac{1-e^{-\lambda}}{\lambda}$ for all $w \in \mathbb{N}$. However, $f_{\mathbb{N}_0}(w) = f_A + f_{A^c}$ solves $\lambda f_{\mathbb{N}_0}(w+1) - w\, f_{\mathbb{N}_0}(w) = 0$, which implies $f_{\mathbb{N}_0}(w) = 0$ for all $w \in \mathbb{N}$. Therefore, $\Delta f_A(w) = -\Delta f_{A^c}(w) \geq \frac{1-e^{-\lambda}}{\lambda}$. We have now proved the following assertion:

Proposition 2.2.1. *For each $A \subseteq \mathbb{N}_0$, there exists a solution f_A to the equation (2.2.5). This solution is uniquely determined on $\mathbb{N}$ and can be arbitrary at 0. Finally, for all $w \in \mathbb{N}$, we have*

$$|\Delta f_A(w)| \leq \frac{1-e^{-\lambda}}{\lambda}.$$

Combining with Proposition 2.1.1, we obtain the following neat result:

Proposition 2.2.2. *For independent random variables $X_i \sim \mathrm{Be}(p_i)$, $i \in \mathscr{I}$, with $\lambda = \sum_{i\in\mathscr{I}} < \infty$ and $W = \sum_{i\in\mathscr{I}} X_i$, we have*

$$\big|\mathbb{P}(W \in A) - \mathrm{Po}(\lambda)(A)\big| \leq \frac{1-e^{-\lambda}}{\lambda} \sum_{i\in\mathscr{I}} p_i^2 \tag{2.2.6}$$

for all $A \subseteq \mathbb{N}_0$.

Remark 2.2.3. We have $\frac{1-e^{-\lambda}}{\lambda} = \int_0^1 e^{-\lambda t}\,\mathrm{d}t \leq 1$ and of course $\frac{1-e^{-\lambda}}{\lambda} \leq \frac{1}{\lambda}$. Estimate (2.2.6) is essentially due to Le Cam [23], who proved the estimates

$$\big|\mathbb{P}(W \in A) - \mathrm{Po}(\lambda)(A)\big| \leq \sum_{i\in\mathscr{I}} p_i^2 \quad \text{and} \quad \big|\mathbb{P}(W \in A) - \mathrm{Po}(\lambda)(A)\big| \leq \frac{8}{\lambda} \sum_{i\in\mathscr{I}} p_i^2$$

(the latter provided that $\max_i p_i \leq 1/4$), using an entirely different argument. However, as we shall see in the sequel, Stein's method can go far beyond independence. This is in contrast to the convolution argument used by Le Cam.

2.3 Abstract results

2.3.1 Coupling

Coupling of two probability distributions means constructing a joint probability distribution with marginals being the two given distributions. In other words, suppose that we have two random variables X and Y, which may be defined on different probability spaces. Coupling means constructing random variables X' and Y' defined on the *same* probability space, such that X' follows the same distribution as X and Y' follows the same distribution as Y.

In applications of Stein's method, coupling is extremely useful. In particular, if W is the random variable whose distribution is to be approximated and X is another random variable related to W, it is often useful to couple the (unconditional) distribution of W with conditional distributions of W given $X = x$. Without loss of generality, one can assume that there are random variables W_x defined on the same probability space as W, such that for each x, the distribution of W_x agrees with the conditional distribution of W given $X = x$.

Example 2.3.1. In Section 1.2, we mentioned that a Stein operator can be constructed by perturbing the original random variable W to a random variable W' with the same

distribution. Now let W, X and W_x be as above and defined on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Suppose that the map $(\omega, \omega') \mapsto W_\omega(\omega')$ is measurable with respect to $\mathscr{F} \otimes \mathscr{F}$. If there is another random variable X' which follows the same distribution of X and is independent of all other random variables, then the random variable $W'(\omega) := W_{X(\omega)}(\omega)$ follows the same distribution as W.

Now let $X_i \sim \mathrm{Be}(p_i)$, $i \in \mathscr{I}$, be again a Bernoulli random variables with $\lambda = \sum_{i\in\mathscr{I}} X_i < \infty$ and $W = \sum_{i\in\mathscr{I}} X_i$, but now we drop the assumption that they are independent. Instead, suppose that for each $i \in \mathscr{I}$, there exists a random variable W_i, such that the (unconditional) distribution of $\tilde{W}_i + 1$ agrees with the conditional distribution of W given $X_i = 1$, or, equivalently, the (unconditional) distribution of $\tilde{W}_i$ agrees with the conditional distribution of $\sum_{j\in\mathscr{I}\setminus\{i\}} X_j$ given $X_i = 1$. Notice that if the random variables X_i are independent, we can simply take $\tilde{W}_i = W_i = W - X_i$. Now observe that

$$\begin{aligned}\mathbb{E}\big[\lambda f(W+1) - f(W)W\big] &= \sum_{i\in\mathscr{I}} \mathbb{E}\big[p_i f(W+1) - f(W)X_i\big] \\ &= \sum_{i\in\mathscr{I}} p_i\, \mathbb{E}\big[f(W+1) - f(\tilde{W}_i+1)\big]\,.\end{aligned}$$

If $|\Delta f(w)| \le M$ for all $w \in \mathbb{N}$, then

$$\big|\mathbb{E}\big[\lambda f(W+1) - f(W)W\big]\big| \le M \sum_{i\in\mathscr{I}} p_i\, \mathbb{E}\,|W - \tilde{W}_i|\,.$$

Combining with Proposition 2.2.1, we find that

$$\big|\mathbb{P}(W \in A) - \mathrm{Po}(\lambda)(A)\big| \le \frac{1-e^{-\lambda}}{\lambda} \sum_{i\in\mathscr{I}} p_i\, \mathbb{E}\,|W - \tilde{W}_i|$$

for all $A \subseteq \mathbb{N}_0$. This can be expressed in terms of the *total variation distance* between two probability measures on $\mathbb{N}_0$ (see Example A.1.2):

$$d_{\mathrm{TV}}(\mu,\nu) := \sup_{A\subseteq\mathbb{N}_0} \big|\mu(A) - \nu(A)\big|\,.$$

Denoting by $\mathscr{L}(W)$ the distribution of W, that is, $\mathscr{L}(W)(A) := \mathbb{P}(W \in A)$, this leads to the following result:

Theorem 2.3.2. *For X_i, W and $\tilde{W}_i$ being as above, we have*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\, \mathrm{Po}(\lambda)\big) \le \frac{1-e^{-\lambda}}{\lambda} \sum_{i\in\mathscr{I}} p_i\, \mathbb{E}\,|W - \tilde{W}_i|\,. \tag{2.3.1}$$

□

Remark 2.3.3. We have only needed to consider conditional distributions of W given $X_i = 1$, not given $X_i = 0$.

Remark 2.3.4. For independent indicators, the preceding theorem reduces to Proposition 2.2.2.

2.3.2 Local dependence and decompositions

Coupling is not the only approach to dependence where Stein's method can be applied. An important alternative concept is *local dependence*. In the context of Stein's method, the latter was introduced by Chen [9] and refined by Arratia, Goldstein and Gordon [1].

Again, take a sum $W = \sum_{i\in\mathscr{I}} X_i$ of Bernoulli random variables $X_i \sim \mathrm{Be}(p_i)$ and assume that for each $i \in \mathscr{I}$, there exists a so called *dependence neighbourhood* $\mathscr{N}_i$, such that X_i is independent of the subfamily X_j, $j \in \mathscr{I} \setminus \big(\{i\} \cup \mathscr{N}_i\big)$. Under this assumption, $\mathscr{N}_i$ can actually be a *punctured neighbourhood* of i.

More generally, suppose that for each i, there is a decomposition

$$W = X_i + Y_i + Z_i\,, \tag{2.3.2}$$

where Z_i is independent of X_i. In this case, write

$$\mathbb{E}\big[\lambda f(W+1) - f(W)W\big] = \sum_{i\in\mathscr{I}} \mathbb{E}\big[p_i f(X_i + Y_i + Z_i + 1) - f(Y_i + Z_i + 1)X_i\big]\,.$$

Since X_i is independent of Z_i, we have

$$\mathbb{E}\big[p_i f(Z_i + 1) - f(Z_i + 1)X_i\big] = 0$$

and therefore

$$\begin{aligned}\mathbb{E}\big[\lambda f(W+1) - f(W)W\big] = \sum_{i\in\mathscr{I}} p_i\,\mathbb{E}\big[f(X_i + Y_i + Z_i + 1) - f(Z_i + 1)\big] \\ - \sum_{i\in\mathscr{I}} \mathbb{E}\big[\big(f(Y_i + Z_i + 1) - f(Z_i + 1)\big)X_i\big]\,.\end{aligned}$$

Now if $|\Delta f(w)| \le M$ for all $w \in \mathbb{N}$, we can estimate

$$\Big|\mathbb{E}\big[\lambda f(W+1) - f(W)W\big]\Big| \le M \sum_{i\in\mathscr{I}} \Big(p_i\,\mathbb{E}\,|X_i + Y_i| + \mathbb{E}\big[X_i\,|Y_i|\big]\Big)\,.$$

Combining with Proposition 2.2.2, we obtain the following result:

Theorem 2.3.5. *For W decomposed as in (2.3.2) and $\tilde{W}_i$ being as above, we have*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\,\mathrm{Po}(\lambda)\big) \le \frac{1-e^{-\lambda}}{\lambda} \sum_{i\in\mathscr{I}} \Big(p_i\,\mathbb{E}\,|X_i + Y_i| + \mathbb{E}\big[X_i\,|Y_i|\big]\Big)\,. \tag{2.3.3}$$

□

Remark 2.3.6. For independent indicators, the preceding theorem again reduces to Proposition 2.2.2.

2.4 Applications

2.4.1 Pattern matching

Let ξ_i, $i \in \mathbb{N}_0$, be a sequence of independent random variables taking values in a finite set L. The elements of L will be called *letters*. Fix a *pattern* $l_0 l_1 \cdots l_{r-1}$. We are interested in the number of occurences of the given pattern in a certain subsequence of the random variables ξ_i. More formally, for $i \in \mathscr{I} := \{0, 1, \ldots, n-1\}$, let

$$X_i := \mathbf{1}(\xi_i = l_0, \xi_{i+1} = l_1, \ldots, \xi_{l+r-1} = l_{r-1})\,.$$

Then $W := \sum_{i=0}^{n-1} X_i$ is the desired number of occurences.

This problem is an obvious case of local dependence, as we can take dependence neighbourhoods

$$\mathscr{N}_i := \{i-r+1, i-r+2, \ldots, i-1, i+1, i+1, \ldots, i+r-1\} \cap \mathscr{I}$$

or, equivalently, take $Y_i := X_{i-r+1} + X_{i-r+2} + \cdots + X_{i-1} + X_{i+1} + X_{i+2} + \cdots + X_{i+r-1}$, letting $X_j = 0$ for $j \notin \mathscr{I}$.

In the sequel, we shall only consider the case where $L = \{0, 1\}$ and patterns of type $11\cdots1$ (so called *r-runs*). In addition, we assume that $\mathbb{P}(\xi_i = 1) = p$ for all i. In this case, we have $\mathbb{E}\, X_i = p^r$ for all i, so that $\lambda = \mathbb{E}\, W = np^r$. Clearly, we have $\mathbb{E}\,|X_i + Y_i| \le (2r-1)p^r$. Noting that

$$\mathbb{E}\big[X_i X_{i+k}\big] = \begin{cases} p^{r+k} & ;\ i, i+k \in \mathscr{I} \\ 0 & ;\ \text{otherwise} \end{cases}$$

for $k = 0, 1, \ldots, r-1$, observe that $\mathbb{E}\big[X_i\,|Y_i|\big] \le 2\sum_{k=1}^{r-1} p^{r+k} = \frac{2(p^{r+1}-p^{2r})}{1-p}$. Summing up, we find that

$$p^r\,\mathbb{E}\,|X_i + Y_i| + \mathbb{E}\big[X_i\,|Y_i|\big] \le (2r-3)p^{2r} + \frac{2p^{r+1}}{1-p}$$

and Theorem 2.3.5 yields

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathscr{L}(W),\, \mathrm{Po}(\lambda)\big) &\le \frac{1-e^{-\lambda}}{\lambda}\, n \left((2r-3)p^{2r} + \frac{2p^{r+1}}{1-p}\right) \\ &\le \left((2r-3)p^r + \frac{2p}{1-p}\right) \min\big\{1, np^r\big\}\,. \end{aligned} \tag{2.4.1}$$

However, it is also easy to construct explicit couplings: observe that a sequence of random letters following the conditional distribution given $X_i = 1$ can be obtained from a sequence following the unconditional distribution simply by setting $\xi_i, \xi_{i+1}, \ldots, \xi_{i+r-1}$ to 1. Denote the new sequence by $\tilde{\xi}_{i,0}, \tilde{\xi}_{i,1}, \ldots$ We may write $\tilde{W}_i = \sum_{j \in \mathscr{I}\setminus\{i\}} \tilde{X}_{ij}$, where $\tilde{X}_{ij} = \mathbf{1}(\tilde{\xi}_{i,j} = \tilde{\xi}_{i,j+1} = \cdots = \tilde{\xi}_{i,j+r-1} = 1)$ (and $\tilde{\xi}_{ij} = 0$ for $j \notin \{0, 1, \ldots, n-1\}$). Clearly, $\tilde{X}_{ij} = X_j$ for $|i-j| \ge r$. Otherwise, observe that $\tilde{X}_{ij}$ differs from X_j only if all appropriate $i-j$ original letters equal 1. Therefore, we have $\mathbb{E}\,|X_j - \tilde{X}_{ij}| \le p^{|i-j|}$ for all $j \ne i$, so that

$$\mathbb{E}\big|W - \tilde{W}_i\big| \le \sum_{j;\,1 \le |i-j| \le r-1} p^{|i-j|} + p^r \le 2\sum_{k=1}^{\infty} p^k \le \frac{2p}{1-p}$$

and Theorem 2.3.2 yields

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\,\mathrm{Po}(\lambda)\big) \le \frac{1-e^{-\lambda}}{\lambda}\,\frac{2np^{r+1}}{1-p} \le \frac{2p}{1-p}\min\big\{1, np^r\big\}\,,$$

which is better than (2.4.1).

2.4.2 Birthday problems

Suppose that there are n people and d days in the year. Each person has his/her birthday on a specific day with probability $1/d$, independently of other persons. Let W be the number of unordered pairs (i. e., sets with exactly two elements) $\{i,j\}$, such that the i-th and the j-th person have the same birthday.

First, we apply the concept of local dependence. Denoting by $\mathscr{I}$ the set of all unordered pairs on $\{1,2,\dots,n\}$, we have $W = \sum_{\{i,j\}\in\mathscr{I}} X_{\{i,j\}}$, where $X_{\{i,j\}}$ is the indicator of the event that the i-th and the j-th person have the same birthday. Letting

$$Y_{\{i,j\}} := \sum_{\substack{1\le k\le n\\ k\ne i,j}} \big(X_{\{i,k\}} + X_{\{k,j\}}\big)\,,$$

observe that $X_{\{i,j\}}$ is independent of $Z_{\{i,j\}} := W - X_{\{i,j\}} - Y_{\{i,j\}}$.

Remark 2.4.1. If i, j and k are distinct, then $X_{\{i,j\}}$ is independent of $X_{\{i,k\}}$ as well as of $X_{\{k,j\}}$. However, it is not *jointly* independent, i. e., it is not independent of the pair $(X_{\{i,k\}}, X_{\{k,j\}})$.

Noting that $\mathbb{E}\,X_{\{i,j\}} = 1/d$ and $\mathbb{E}\big[X_{\{i,j\}}X_{\{i,k\}}\big] = \mathbb{E}\big[X_{\{i,j\}}X_{\{k,j\}}\big] = 1/d^2$ for all distinct i,j,k, we have $\lambda = \mathbb{E}\,W = \frac{n(n-1)}{2d}$ and Theorem 2.3.5 yields

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathscr{L}(W),\,\mathrm{Po}(\lambda)\big) &\le \frac{1-e^{-\lambda}}{\lambda}\,\frac{n(n-1)(4n-7)}{2d^2} \\ &\le \min\left\{\frac{4n-7}{d},\, \frac{n(n-1)(4n-7)}{2d^2}\right\}. \end{aligned} \tag{2.4.2}$$

Alternatively, one can also construct explicit couplings. Fixing the i-th and the j-th person, observe that a random pair of birthdays following the conditional distribution given that they both have the same birthday can be obtained from a pair following the unconditional distribution by picking one of them independently and uniformly at random and assigning the other the birthday of the chosen one. Since the birthdays are independent, we can leave the other persons unchanged. Accordingly, we construct $\tilde{W}_{\{i,j\}}$. We may write $\tilde{W}_{\{i,j\}} = \sum_{\{k,l\}\in\mathscr{I}\setminus\{\{i,j\}\}} \tilde{X}_{\{i,j\},\{k,l\}}$, where $\tilde{X}_{\{i,j\},\{k,l\}}$ is the indicator of the event that k and l have the same birthday after the above-mentioned reassignment. Clearly, $\tilde{X}_{\{i,j\},\{k,l\}} = X_{\{k,l\}}$ if $\{i,j\}\cap\{k,l\} = \emptyset$. Otherwise, we may assume without loss of generality that $k = i$ (and $l \ne i,j$). In this case, $\tilde{X}_{\{i,j\},\{i,l\}}$ differs from $X_{\{i,j\}}$ if, first, i

is to be assigned the birthday of j, and second, if among the pairs $\{i,l\}$ and $\{j,l\}$, there is exactly one of them such that both persons have the same birthday. Therefore,

$$\mathbb{E}\big|X_{\{i,j\}} - \tilde{X}_{\{i,j\},\{i,l\}}\big| = \frac{d-1}{d^2}$$

and

$$\mathbb{E}\big|W - \tilde{W}_{\{i,j\}}\big| \le \frac{2(n-2)(d-1)}{d^2} + \frac{1}{d} = \frac{2nd-3d-2n+4}{d^2}\,.$$

COmpared to (2.4.2), Theorem 2.3.2 yields a better bound

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\,\mathrm{Po}(\lambda)\big) \le \frac{1-e^{-\lambda}}{\lambda}\,\frac{n(n-1)(2nd-3d-2n+4)}{2d^3}\,. \tag{2.4.3}$$

The exact probability $\mathbb{P}(W=0)$ is easy to compute – we have

$$\mathbb{P}(W=0) = \frac{(d-1)(d-2)\cdots(d-n+1)}{d^{n-1}}$$

and for $d=365, n=23$, we have $\mathbb{P}(W=0) \doteq 0.4927028$ (closest to $1/2$ for $d=365$). Poisson approximation yields $\mathbb{P}(W=0) \approx 0.4999982$. From (2.4.3), we obtain an error bound 0.0587, which is quite larger than the actual error.

Remark 2.4.2. This example has many possible extensions. Among others, the birthday probabilities need not be equal and we may only consider pairs which form an edge in a certain given graph (i. e., we consider coloured graphs). Both extensions can be handled by Stein's method, see Barbour, Holst and Janson [5].

2.4.3 Random permutations

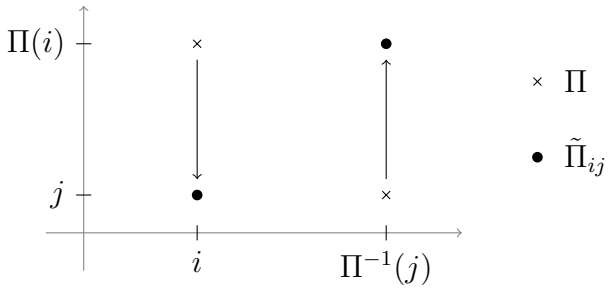
Let Π be a uniformly distributed random permutation of $\{1,2,\dots,n\}$. Take $C \subseteq \{1,2,\dots,n\}^2$ and consider the statistic

$$W := \sum_{i=1}^{n} \mathbf{1}\big((i,\Pi(i) \in C\big)\,.$$

The indicators $\mathbf{1}\big((i,\Pi(i)) \in C\big)$ do not exhibit any useful local dependence: any two of them are dependent. However, it is easy to construct a proper coupling along with a bound on the error in the Poisson approximation. One can rewrite W as

$$W = \sum_{(i,j)\in C} X_{ij}\,, \tag{2.4.4}$$

where $X_{ij} := \mathbf{1}\big(\Pi(i)=j\big)$. Given $\Pi(i)=j$, Π is uniformly distributed over all permutations $\tilde{\pi}$ with $\tilde{\pi}(i)=j$. Now define a new random permutation $\tilde{\Pi}_{ij} := \tau_{\Pi(i),j} \circ \Pi$, where $\tau_{r,s}$ is the transposition exchanging r and s (and identity if $r=s$):



We claim that $\tilde{\Pi}_{ij}$ is uniformly distributed over all permutations $\tilde{\pi}$ with $\tilde{\pi}(i) = j$. If π is a fixed permutation and $\tilde{\pi} = \tau_{\pi(i),j} \circ \pi$, then, clearly, $\tilde{\pi}(i) = j$. Now fix a permutation $\tilde{\pi}$ with $\tilde{\pi}(i) = j$ and examine all permutations π with $\tilde{\pi} = \tau_{\pi(i),j} \circ \pi$. As $\pi = \tau_{j,\pi(i)} \circ \tilde{\pi}$, π must be of the form $\pi = \tau_{j,l} \circ \tilde{\pi}$ for some $l \in \{1, 2, \dots, n\}$. Conversely, if $\pi = \tau_{j,l} \circ \tilde{\pi}$, then $\pi(i) = l$ and $\tau_{\pi(i),j} \circ \pi = \tau_{l,j} \circ \tau_{j,l} \circ \tilde{\pi} = \tilde{\pi}$. Therefore, for each $\tilde{\pi}$ with $\tilde{\pi}(i) = j$, there are exactly n permutations π, such that $\tilde{\pi} = \tau_{\pi(i),j} \circ \pi$. Therefore, $\tilde{\Pi}_{ij}$ is indeed uniformly distributed over all permutations $\tilde{\pi}$ with $\tilde{\pi}(i) = j$. Letting $\tilde{X}_{ijkl} := \mathbf{1}\big(\tilde{\Pi}_{ij}(k) = l\big)$ and

$$\tilde{W}_{ij} = \sum_{(k,l)\in C} \tilde{X}_{ijkl} - 1\,,$$

the (unconditional) distribution $\tilde{W}+1$ agrees with the conditional distribution of W given $X_{ij} = 1$.

Now observe that $\tilde{\Pi}_{ij}$ agrees with Π except on i and $\Pi^{-1}(j)$. More precisely, if $\Pi(i) = j$, then $\tilde{X}_{ijkl} = X_{kl}$ for all $(k,l) \in C$. If $\Pi(i) \neq j$, then $\tilde{X}_{ijkl}$ differs from X_{kl} in the following four disjoint cases:

- If $k = i$ and $l = j$, then $X_{kl} = 0$ and $\tilde{X}_{ijkl} = 1$.
- If $k = i$ and $l = \Pi(i)$, then $X_{kl} = 1$ and $\tilde{X}_{ijkl} = 0$.
- If $\Pi(k) = j$ and $l = j$, then $X_{kl} = 1$ and $\tilde{X}_{ijkl} = 0$.
- If $\Pi(k) = j$ and $l = \Pi(i)$, then $X_{kl} = 0$ and $\tilde{X}_{ijkl} = 1$.

Therefore, if $(i,j) \in C$, then

$$\begin{aligned} W - \tilde{W}_{ij} &= \sum_{(k,l)\in C} \big(X_{kl} - \tilde{X}_{ijkl}\big) + 1 \\ &= \sum_{l;(i,l)\in C} \mathbf{1}\big(\Pi(i) = l\big) + \sum_{k;(k,j)\in C} \mathbf{1}\big(\Pi(k) = j\big) - \sum_{(k,l)\in C} \mathbf{1}\big(\Pi(i) = l,\, \Pi(k) = j\big)\,. \end{aligned}$$

Next, observe that the preceding identity remains true if $\Pi(i) = j$. Noting that the event $\{\Pi(i) = l,\, \Pi(k) = j\}$ is only possible if either $k = i$ and $l = j$ or $k \neq i$ and $l \neq j$, observe that

$$\begin{aligned} W - \tilde{W}_{ij} = \mathbf{1}\big(\Pi(i) = j\big) &+ \sum_{\substack{l;(i,l)\in C \\ l\neq j}} \mathbf{1}\big(\Pi(i) = l\big) + \sum_{\substack{k;(k,j)\in C \\ k\neq i}} \mathbf{1}\big(\Pi(k) = j\big) \\ &- \sum_{\substack{(k,l)\in C \\ k\neq i, l\neq j}} \mathbf{1}\big(\Pi(i) = l,\, \Pi(k) = j\big)\,. \end{aligned}$$

Noting that $\mathbb{P}(\Pi(i)=j)=\mathbb{P}(\Pi(k)=j)=\frac{1}{n}$ and $\mathbb{P}(\Pi(i)=l,\,\Pi(k)=j)=\frac{1}{n(n-1)}$ for $k\neq i, l\neq j$, we estimate

$$\begin{aligned}\mathbb{E}\big|W-\tilde{W}_{ij}\big| &\leq \frac{1}{n}+\frac{1}{n}\big|\big\{l\,;\,(i,l)\in C, l\neq j\big\}\big|+\frac{1}{n}\big|\big\{k\,;\,(k,j)\in C, k\neq i\big\}\big| \\ &\qquad +\frac{1}{n(n-1)}\big|\big\{(k,l)\in C\,;\,k\neq i, l\neq j\big\}\big| \\ &= \frac{n-2}{n(n-1)}+\frac{n-2}{n(n-1)}\big|\big\{l\,;\,(i,l)\in C, l\neq j\big\}\big| \\ &\qquad +\frac{n-2}{n(n-1)}\big|\big\{k\,;\,(k,j)\in C, k\neq i\big\}\big|+\frac{1}{n(n-1)}|C| \\ &= -\frac{n-2}{n(n-1)}+\frac{n-2}{n(n-1)}\big|\big\{l\,;\,(i,l)\in C\big\}\big| \\ &\qquad +\frac{n-2}{n(n-1)}\big|\big\{k\,;\,(k,j)\in C\big\}\big|+\frac{1}{n(n-1)}|C|\end{aligned}$$

For each $(i,j)\in C$, we have $\mathbb{P}(X_{ij}=1)=\frac{1}{n}$. Summing up, we obtain

$$\begin{aligned}&\sum_{(i,j)\in C}\mathbb{P}(X_{ij}=1)\,\mathbb{E}\big|W-\tilde{W}_{ij}\big| \\ &\qquad\leq -\frac{n-2}{n^2(n-1)}|C|+\frac{n-2}{n^2(n-1)}\big|\big\{(i,j,l)\,;\,(i,j)\in C,(i,l)\in C\big\}\big| \\ &\qquad\qquad +\frac{n-2}{n^2(n-1)}\big|\big\{(i,k,l)\in C\,;\,(i,j)\in C,(k,j)\in C\big\}\big| \\ &\qquad\qquad +\frac{1}{n^2(n-1)}|C|^2\,.\end{aligned}$$

Introducing

$$p_i:=\frac{\big|\big\{j\,;\,(i,j)\in C\big\}\big|}{n}\,,\qquad q_j:=\frac{\big|\big\{j\,;\,(i,j)\in C\big\}\big|}{n}\,,\qquad \lambda:=\frac{|C|}{n}=\mathbb{E}\,W \tag{2.4.5}$$

and applying Theorem 2.3.2, we obtain

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\mathrm{Po}(\lambda)\big)\leq\frac{1-e^{-\lambda}}{\lambda}\left(\frac{n-2}{n-1}\sum_{i=1}^n p_i^2+\frac{n-2}{n-1}\sum_{j=1}^n q_j^2+\frac{\lambda^2}{n-1}-\frac{(n-2)\lambda}{n(n-1)}\right). \tag{2.4.6}$$

Remark 2.4.3. If a constant factor in the error bound does not matter, the bound in (2.4.6) can be simplified. First, by the inequality between the arithmetic and geometric mean, we can estimate

$$\lambda^2=\sum_{i=1}^n p_i\sum_{j=1}^n q_j\leq\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n(p_i^2+q_j^2)=\frac{n}{2}\left(\sum_{i=1}^n p_i^2+\sum_{j=1}^n q_j^2\right),$$

leading to

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\,\mathrm{Po}(\lambda)\big)\leq\frac{1-e^{-\lambda}}{\lambda}\left(\frac{3n-4}{2(n-1)}\sum_{i=1}^n p_i^2+\frac{3n-4}{2(n-1)}\sum_{j=1}^n q_j^2-\frac{(n-2)\lambda}{n(n-1)}\right).$$

In addition, since $p_i \le np_i^2$, we can estimate

$$-\frac{(n-2)\lambda}{n(n-1)} = -\frac{\lambda}{n} + \frac{1}{n(n-1)}\sum_{i=1}^{n} p_i \le -\frac{\lambda}{n} + \frac{1}{n-1}\sum_{i=1}^{n} p_i^2$$

combining all together, we obtain a simplified bound

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\, \mathrm{Po}(\lambda)\big) \le \frac{1-e^{-\lambda}}{\lambda}\left(\frac{3}{2}\sum_{i=1}^{n} p_i^2 + \frac{3}{2}\sum_{j=1}^{n} q_j^2 - \frac{\lambda}{n}\right).$$

Below we give two examples.

Example 2.4.4. The case of independent indicators can be regarded as a limiting case of random permutations. For the sake of simplicity, we assume that we have finitely many independent indicators $X_1, X_2, \dots, X_r$ with success probabilities $p_1, p_2, \dots, p_r$ all divisible by the same number $s \in \mathbb{N}$. For each $m \in \mathbb{N}$, define

$$C^{(m)} := \bigcup_{i=1}^{r} \{i\} \times \{1, 2, \dots, mp_i s\}\,.$$

For $n := ms \ge r$, let $\Pi^{(m)}$ be a uniformly distributed random permutation of $\{1, 2, \dots, n\}$. Define

$$W^{(m)} := \sum_{i=1}^{n} X_i^{(m)} = \sum_{i=1}^{r} X_i^{(m)}\,,$$

where $X_i^{(m)} = \mathbf{1}\big((i, \Pi^{(m)}(i)) \in C\big)$. Observe that $X_i^{(m)} \sim \mathrm{Be}(p_i)$. It is plausible that in the limit as m tends to the infinity, the random variables $X_1^{(m)}, \dots, X_r^{(m)}$ are independent. Next, observe that p_i and λ match the underlying quantities defined in (2.4.5). Letting $q_j^{(m)}$ be the counterpart of q_j in (2.4.5), observe that $q_j^{(m)} \le r/n$. Therefore, by (2.4.6), we have

$$d_{\mathrm{TV}}\Big(\mathscr{L}\big(W^{(m)}\big),\, \mathrm{Po}(\lambda)\Big) \le \frac{1-e^{-\lambda}}{\lambda}\left(\frac{n-2}{n-1}\sum_{i=1}^{n} p_i^2 + \frac{n-2}{n-1}\,\frac{r^2}{n} + \frac{\lambda^2}{n-1} - \frac{(n-2)\lambda}{n(n-1)}\right)$$

and the right hand side tends to the bound in (2.2.6).

Example 2.4.5. **Matching problem**. Consider the setting of $n = rd$ cards, d if which have face value i, $i = 1, 2, \dots, r$, and draw r of them uniformly at random. A match occurs if a card with face i appears in the i-th drawing. Denoting by W the number of matchings, observe that this can be represented in form (2.4.4) if we take

$$C := \bigcup_{i=1}^{r} \{i\} \times \{d(i-1)+1, d(i-1)+2, \dots, d(i-1)+d\}\,.$$

Next, we have

$$p_i = \frac{1}{r}\,\mathbf{1}(i \le r)\,, \qquad q_j = \frac{1}{n}\,, \qquad \lambda = 1$$

and the bound (2.4.6) reduces after some calculation to

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\, \mathrm{Po}(\lambda)\big) \le \big(1 - e^{-1}\big) \left(\frac{n-2}{n-1} \frac{1}{r} + \frac{1}{n-1} \right) .$$

A matching problem was investigated in 1708 in de Montmort's book [14] on games of chance for the game 'Treize', where we have an ordinary deck of $n = 52$ cards taking $r = 13$ possible face values. The original problem was to find the probability that there was no match. The approximating probability equals $\mathrm{Po}(1)(\{0\}) = 1/e \doteq 0.3679$. The exact probability can be computed by the inclusion–exclusion principle and equals 0.3569 (and was computed in 1711 by Nicolas Bernoulli, see page 324 of de Montmort [14]). The difference equals 0.0109, while the error bound equals 0.06 (all up to rounding errors).

2.4.4 Occupancy problems

Let r balls be thrown independently into a family of bins indexed by $\mathscr{I}$, with probability q_i of hitting the i-th bin. Suppose first that $\mathscr{I}$ is finite and define W to be the number of empty bins. We can write $W = \sum_{i \in \mathscr{I}} X_i$, where X_i is the indicator of the event that the i-th bin is empty. Clearly, $X_i \sim \mathrm{Be}(p_i)$, where

$$p_i = (1 - q_i)^r .$$

The event $\{X_i = 1\}$ is the same as the event that the balls have been only thrown into bins in $\mathscr{I} \setminus \{i\}$. Conditioning on this event is the same as modifying the probability q_j to $q_j/(1 - q_i)$ for $j \ne i$ and q_i to 0. This is the same as first throwing the balls as initially and then relocating each ball that has landed in the i-th bin into the j-th bin, $j \ne i$, with probability $q_j/(1 - q_i)$. Taking W to be the number of empty bins before the rearrangement and $\tilde{W}_i$ the number of empty bins except for the i-th bin after the rearrangement, we obtain the desired coupling.

We can write $\tilde{W}_i = \sum_{j \in \mathscr{I} \setminus \{i\}} \tilde{X}_{ij}$, where $\tilde{X}_{ij}$ is the indicator of the event that the j-th bin is empty after the rearrangement. Write

$$W - \tilde{W}_i = X_i + \sum_{j \in \mathscr{I} \setminus \{i\}} (X_j - \tilde{X}_{ij}) .$$

If the j-th bin is empty after the rearrangement, it must have been empty before the rearrangement, too. Therefore, $X_j \ge \tilde{X}_{ij}$ and

$$\begin{aligned} \mathbb{E}\,|W - \tilde{W}_i| &= \mathbb{E}\, X_i + \sum_{j \in \mathscr{I} \setminus \{i\}} (\mathbb{E}\, X_j - \mathbb{E}\, \tilde{X}_{ij}) \\ &= (1 - q_i)^r + \sum_{j \in \mathscr{I} \setminus \{i\}} \left[(1 - q_j)^r - \left(1 - \frac{q_j}{1 - q_i} \right)^r \right] . \end{aligned}$$

We can further estimate this quantity by rewriting

$$(1-q_j)^r - \left(1-\frac{q_j}{1-q_i}\right)^r = (1-q_j)^r - \left(1-q_j-\frac{q_iq_j}{1-q_i}\right)^r$$
$$= (1-q_j)^r\left[1-\left(1-\frac{q_iq_j}{(1-q_i)(1-q_j)}\right)^r\right].$$

Applying the inequality $(1-x)^r \geq 1-rx$, which holds true for all $x \in [0,1]$, estimate

$$(1-q_j)^r - \left(1-\frac{q_j}{1-q_i}\right)^r \leq r(1-q_j)^r\,\frac{q_iq_j}{(1-q_i)(1-q_j)} = \frac{rp_jq_iq_j}{(1-q_i)(1-q_j)}.$$

Letting $\lambda = \sum_{i\in\mathscr{I}} p_i$, combining all together and applying Theorem 2.3.2, we obtain the following result:

Proposition 2.4.6. *The distribution of the number W of empty bins defined as above satisfies*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\,\mathrm{Po}(\lambda)\big) \leq \frac{1-e^{-\lambda}}{\lambda}\sum_{i\in\mathscr{I}}\left(p_i^2 + r\sum_{j\in\mathscr{I}\setminus\{i\}}\frac{p_ip_jq_iq_j}{(1-q_i)(1-q_j)}\right). \tag{2.4.7}$$

□

Now suppose that all bins are hit with equal probabilities, that is, $q_i = \frac{1}{n}$, where n denotes the number of bins. In this case, (2.4.7) reduces to

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\,\mathrm{Po}(\lambda)\big) \leq n\,\frac{1-e^{-\lambda}}{\lambda}\left(1-\frac{1}{n}\right)^{2r}\left(1+\frac{r}{n-1}\right),$$

where $\lambda = n\big(1-\frac{1}{n}\big)^r$. Letting $a := r/n$, observe that $p_i \leq e^{-a}$ and consequently $\lambda \leq n\,e^{-a}$. If a is not too large, the latter bound is of correct order. In this case, by Remark 2.2.3, $\frac{1-e^{-\lambda}}{\lambda}$ is of order $\min\big\{1, \frac{1}{n}\,e^a\big\}$. As it turns out, this case is an upper bound.

Lemma 2.4.7. *There exists a constant B, such that, letting $\lambda = n\big(1-\frac{1}{n}\big)^{an}$, we have $\frac{1-e^{-\lambda}}{\lambda} \leq \min\big\{1, \frac{B\,e^a}{n}\big\}$ for all $n \geq 2$ and $a \geq 0$.*

Corollary 2.4.8. *Let W be as above. There exists a constant C, such that*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\,\mathrm{Po}(\lambda)\big) \leq C\Big((1+a)\,e^{-a}\min\big\{1, n\,e^{-a}\big\}\Big).$$

for all $n \geq 2$ and $r \geq 0$, where $a = r/n$. In particular, the total variation error in the Poisson approximation tends to zero uniformly in n as $a \to \infty$. □

Proof of Lemma 2.4.7. Applying Taylor expansion, we find that

$$an\log\left(1-\frac{1}{n}\right) = -a - \frac{a}{2n\big(1-\frac{\theta}{n}\big)^2} \geq -a - \frac{2a}{n} \tag{2.4.8}$$

for some $\theta \in [0,1]$. Therefore, for $a \leq n/2$, we have $\lambda \geq n\,e^{-a-1}$. By Remark 2.2.3, it follows that $\frac{1-e^{-\lambda}}{\lambda} \leq \min\big\{1, \frac{1}{\lambda}\big\} \leq \min\big\{1, \frac{e^{a+1}}{n}\big\}$. On the other hand, for $a \geq n/2$, observe that $\frac{e^{a+1}}{n} \geq \frac{e^{a+1}}{2a} \geq \frac{e^2}{2} \geq 1$, so that $\frac{1-e^{-\lambda}}{\lambda} \leq 1 = \min\big\{1, \frac{e^{a+1}}{n}\big\}$. This proves the result with $B = e$. □

Now consider the more general case considering the number of bins containing exactly m balls; denote this number by $W^{(m)}$. Again, $W^{(m)} = \sum_{i\in\mathscr{I}} X_i^{(m)}$, where $X_i^{(m)}$ is the indicator of the event that the i-th bin contains exactly m balls. Clearly, $X_i^{(m)} \sim \mathrm{Be}\big(p_i^{(m)}\big)$, where

$$p_i^{(m)} = \binom{r}{m} q_i^m (1-q_i)^{r-m}\,.$$

We have already constructed a coupling of the (unconditional) distribution of W with its conditional distribution given that the i-th bin is empty. Continuing from the balls rearranged in the latter way, further pick m balls uniformly at random and relocate them into the i-th bin. This makes a coupling with the conditional distribution given $X_i^{(m)} = 1$.

Remark 2.4.9. Notice that the latter coupling is not entirely optimal. A more refined coupling would go as follows: consider the number of balls in the i-th bin. If there are exactly m balls, leave it as it is. If there are more, suitably relocate the excess balls into other bins. If they are less, pick a suitable number of balls from other bins uniformly at random and relocate them into the i-th bin. However, in this case, the right hand side of (2.3.1) is more complicated to estimate and we do not benefit on the rate of convergence. Therefore, we keep the afore-mentioned two-step coupling.

Let $\tilde{X}_{ij}^{(m)}$ denote the indicator of the event that the j-th bin contains exactly m balls after relocating all balls from the i-th bin. Next, let $\tilde{\tilde{X}}_{ij}^{(m)}$ denote the indicator of the same event after relocating m balls into the i-th bin. As usual, let $\tilde{W}_i^{(m)} := \sum_{j\in\mathscr{I}\setminus\{i\}} \tilde{X}_{ij}^{(m)}$ and $\tilde{\tilde{W}}_i^{(m)} := \sum_{j\in\mathscr{I}\setminus\{i\}} \tilde{\tilde{X}}_{ij}^{(m)}$. The (unconditional) distribution of $1+\tilde{\tilde{W}}_i^{(m)}$ agrees with the conditional distribution of W given $X_i = 1$.

Now estimate

$$\begin{aligned}
\big|W^{(m)} - \tilde{\tilde{W}}_i^{(m)}\big| &\le \big|W^{(m)} - \tilde{W}_i^{(m)}\big| + \big|\tilde{W}_i^{(m)} - \tilde{\tilde{W}}_i^{(m)}\big| \\
&\le X_i + \sum_{j\in\mathscr{I}\setminus\{i\}} \Big[\big|X_j^{(m)} - \tilde{X}_{ij}^{(m)}\big| + \big|\tilde{X}_{ij}^{(m)} - \tilde{\tilde{X}}_{ij}^{(m)}\big|\Big] \\
&= X_i + \sum_{j\in\mathscr{I}\setminus\{i\}} \Big[\big(X_j^{(m)} - X_j^{(m)}\tilde{X}_{ij}^{(m)}\big) + \big(\tilde{X}_{ij}^{(m)} - X_j^{(m)}\tilde{X}_{ij}^{(m)}\big) \\
&\qquad + \big(\tilde{X}_{ij}^{(m)} - \tilde{X}_{ij}^{(m)}\tilde{\tilde{X}}_{ij}^{(m)}\big) + \big(\tilde{\tilde{X}}_{ij}^{(m)} - \tilde{X}_{ij}^{(m)}\tilde{\tilde{X}}_{ij}^{(m)}\big)\Big]
\end{aligned}$$

(notice that $X_j^{(m)} \ge X_j^{(m)}\tilde{X}_{ij}^{(m)}$ and similarly for the other differences).

Now compute the expectations of all random variables, i. e., the probabilities of all underlying events.

- Recall that $\mathbb{E}\, X_j^{(m)} = \mathbb{E}\, p_i^{(m)} = \binom{r}{m} q_i^m (1-q_i)^{r-m}$.
- After the first rearrangement, each ball is in the j-th bin with probability $q_j/(1-q_i)$, where the balls are independent. Therefore,

$$\mathbb{E}\,\tilde{X}_{ij}^{(m)} = \binom{r}{m} \left(\frac{q_j}{1-q_i}\right)^m \left(1 - \frac{q_j}{1-q_i}\right)^{r-m}\,.$$

- In the second rearrangement, m balls are selected to be relocated to the i-th bin. Given the choice of these m balls, each one of the rest is in the j-th bin with probability $q_j/(1-q_i)$, where the balls are again independent. Therefore,

$$\mathbb{E}\,\tilde{\tilde{X}}_{ij}^{(m)} = \binom{r-m}{m}\left(\frac{q_j}{1-q_i}\right)^m\left(1-\frac{q_j}{1-q_i}\right)^{r-2m}.$$

- The product $X_j^{(m)}\tilde{X}_{ij}^{(m)}$ is the indicator of the event that in the j-th bin, there are exactly m balls before and after the first rearrangement. There are $\binom{r}{m}$ choices of balls that first land in the j-th bin, and for each of them, this occurs with probability q_j. Given the choice of these m balls, each one of the remaining $r-m$ balls lands in the j-th bin with probability $q_j/(1-q_i)$ after the first rearrangement, and the balls are independent. Therefore,

$$\mathbb{E}\big[X_j^{(m)}\tilde{\tilde{X}}_{ij}^{(m)}\big] = \binom{r}{m}q_j^m\left(1-\frac{q_j}{1-q_i}\right)^{r-m}.$$

- The product $\tilde{X}_{ij}^{(m)}\tilde{\tilde{X}}_{ij}^{(m)}$ is the indicator of the event that in the j-th bin, there are exactly m balls before and after the second rearrangement. Recall that in the second rearrangement, m balls are selected to be relocated to the i-th bin. We can choose these balls first. None of these balls should be in the j-th bin after the first rearrangement. Given the selection of these m balls, exactly m among the remaining $r-m$ balls should be in the j-th bin after the first (and the second) rearrangement. Since the selection of the balls to be relocated in the second turn is independent of the location of the balls after the first rearrangement, we find that

$$\mathbb{E}\big[\tilde{X}_{ij}^{(m)}\tilde{\tilde{X}}_{ij}^{(m)}\big] = \binom{r-m}{m}\left(\frac{q_j}{1-q_i}\right)^m\left(1-\frac{q_j}{1-q_i}\right)^{r-m}.$$

Now estimate

$$\begin{aligned}
\mathbb{E}\big(X_j^{(m)} - X_j^{(m)}\tilde{X}_{ij}^{(m)}\big) &= \binom{r}{m}q_j^m\left[(1-q_j)^{r-m} - \left(1-\frac{q_j}{1-q_i}\right)^{r-m}\right] \\
&= \binom{r}{m}q_j^m\left[(1-q_j)^{r-m} - \left(1-q_j-\frac{q_iq_j}{1-q_i}\right)^{r-m}\right] \\
&= \binom{r}{m}q_j^m(1-q_j)^{r-m}\left[1-\left(1-\frac{q_iq_j}{(1-q_i)(1-q_j)}\right)^{r-m}\right] \\
&\le (r-m)\,\frac{q_iq_j}{(1-q_i)(1-q_j)}\binom{r}{m}q_j^m(1-q_j)^{r-m},
\end{aligned}$$

$$\begin{aligned}\mathbb{E}\big(\tilde{X}_{ij}^{(m)} - X_j^{(m)}\tilde{X}_{ij}^{(m)}\big) &= \binom{r}{m}\left(1-\frac{q_j}{1-q_i}\right)^{r-m}\left[\left(\frac{q_j}{1-q_i}\right)^m - q_j^m\right]\\ &= \binom{r}{m}\left(\frac{q_j}{1-q_i}\right)^m\left(1-\frac{q_j}{1-q_i}\right)^{r-m}\left[1-(1-q_i)^m\right]\\ &\le mq_i\binom{r}{m}\left(\frac{q_j}{1-q_i}\right)^m\left(1-\frac{q_j}{1-q_i}\right)^{r-m},\end{aligned}$$

$$\begin{aligned}\mathbb{E}\big(\tilde{X}_{ij}^{(m)} - \tilde{X}_{ij}^{(m)}\tilde{\tilde{X}}_{ij}^{(m)}\big) &= \left[\binom{r}{m} - \binom{r-m}{m}\right]\left(\frac{q_j}{1-q_i}\right)^m\left(1-\frac{q_j}{1-q_i}\right)^{r-m}\\ &= \sum_{k=r-m}^{r-1}\binom{k}{m-1}\left(\frac{q_j}{1-q_i}\right)^m\left(1-\frac{q_j}{1-q_i}\right)^{r-m}\\ &\le m\binom{r-1}{m-1}\left(\frac{q_j}{1-q_i}\right)^m\left(1-\frac{q_j}{1-q_i}\right)^{r-m}\\ &= \frac{m^2}{r}\binom{r}{m}\left(\frac{q_j}{1-q_i}\right)^m\left(1-\frac{q_j}{1-q_i}\right)^{r-m}.\end{aligned}$$

$$\begin{aligned}\mathbb{E}\big(\tilde{\tilde{X}}_{ij}^{(m)} - \tilde{X}_{ij}^{(m)}\tilde{\tilde{X}}_{ij}^{(m)}\big) &= \binom{r-m}{m}\left(\frac{q_j}{1-q_i}\right)^m\left(1-\frac{q_j}{1-q_i}\right)^{r-2m}\left[1-\left(1-\frac{q_j}{1-q_i}\right)^m\right]\\ &\le m\,\frac{q_j}{1-q_i}\binom{r-m}{m}\left(\frac{q_j}{1-q_i}\right)^m\left(1-\frac{q_j}{1-q_i}\right)^{r-2m}.\end{aligned}$$

Collecting all together, we obtain (provided that $m \ge 1$)

$$\mathbb{E}\big|W^{(m)} - \tilde{\tilde{W}}_i^{(m)}\big| \le p_i^{(m)} + \sum_{j\in\mathscr{I}\setminus\{i\}} p_{ij}^{(m)}\left[\frac{(r-m)q_iq_j}{1-q_j} + mq_i + m\,\frac{q_j}{1-q_i} + \frac{m^2}{r}\right],$$

where

$$p_{ij}^{(m)} := \binom{r}{m}\left(\frac{q_j}{1-q_i}\right)^m(1-q_j)^{r-2m}.$$

Letting $\lambda_m = \sum_{i\in\mathscr{I}} p_i^{(m)}$ and applying Theorem 2.3.2, we obtain the following result:

Proposition 2.4.10. *For $m \ge 1$, the distribution of the number $W^{(m)}$ of the bins with exactly m balls satisfies*

$$\begin{aligned}&d_{\mathrm{TV}}\Big(\mathscr{L}\big(W^{(m)}\big),\,\mathrm{Po}(\lambda_m)\Big)\\ &\quad\le \frac{1-e^{-\lambda_m}}{\lambda_m}\sum_{i\in\mathscr{I}}\Bigg(\big(p_i^{(m)}\big)^2\\ &\qquad + \sum_{j\in\mathscr{I}\setminus\{i\}} p_i^{(m)}p_{ij}^{(m)}\left[\frac{(r-m)q_iq_j}{1-q_j} + mq_i + m\,\frac{q_j}{1-q_i} + \frac{m^2}{r}\right]\Bigg).\end{aligned} \tag{2.4.9}$$

□

Now consider again the case where all bins are hit with equal probabilities, that is, $q_i = \frac{1}{n}$, where n denotes the number of bins. In this case, (2.4.9) reduces to

$$d_{\mathrm{TV}}\Big(\mathscr{L}\big(W^{(m)}\big),\, \mathrm{Po}(\lambda_m)\Big) \le \frac{1-e^{-\lambda_m}}{\lambda_m}\Bigg[n\,\big(p_1^{(m)}\big)^2 + \left(r + 2m(n-1) + \frac{m^2 n(n-1)}{r}\right) p_1^{(m)} p_{12}^{(m)}\Bigg],$$

where

$$p_1^{(m)} = \binom{r}{m}\frac{1}{n^m}\left(1-\frac{1}{n}\right)^{r-m}, \quad p_{12}^{(m)} = \binom{r}{m}\frac{1}{(n-1)^m}\left(1-\frac{1}{n}\right)^{r-2m}, \quad \lambda_m = n\,p_1^{(m)}.$$

Suppose that $n \ge 2$ and $r \ge m$. Letting $a := r/n$, observe that

$$p_i^{(m)} \le \frac{r^m}{m!}\frac{1}{n^m}\,e^{-a}\left(\frac{n}{n-1}\right)^m \le \frac{(2a)^m}{m!}\,e^{-a}$$

and

$$p_{ij}^{(m)} \le \frac{r^m}{m!}\frac{1}{n^m}\,e^{-a}\left(\frac{n}{n-1}\right)^{3m} \le \frac{(8a)^m}{m!}\,e^{-a}$$

Thus, λ is of order at most $n\,a^m e^{-a}$. If this is the actual order, then, by Remark 2.2.3, $\frac{1-e^{-\lambda}}{\lambda}$ is of order $\min\{1, \frac{e^a}{n\,a^m}\}$. Again, this order is an upper bound, as shown in the following generalization of Lemma 2.4.7:

Lemma 2.4.11. *For each $m \in \mathbb{N}_0$, there exists a constant B_m, such that, letting $\lambda = \binom{an}{m}\frac{1}{n^{m-1}}\big(1-\frac{1}{n}\big)^{an-m}$, we have $\frac{1-e^{-\lambda}}{\lambda} \le \min\{1, \frac{B_m\,e^a}{n\,a^m}\}$ for all $n \ge 2$ and $a > 0$.*

Corollary 2.4.12. *Let $W^{(m)}$ and λ_m be as above. For each $m \in \mathbb{N}_0$, there exists a constant C_m, such that*

$$d_{\mathrm{TV}}\Big(\mathscr{L}\big(W^{(m)}\big),\, \mathrm{Po}(\lambda_m)\Big) \le C_m\Big(\big(a^{m+1}+a^{m-1}\big)\,e^{-a}\min\big\{1, n\,a^m e^{-a}\big\}\Big).$$

for all $n \ge 2$ and $r \ge 0$, where $a = r/n$. In particular, the total variation error in the Poisson approximation tends to zero uniformly in n as $a \to \infty$. For $m \ge 2$, this is also true as $a \to 0$. □

Remark 2.4.13. For $m = 1$, the error does not tend to zero uniformly in n as $a \to 0$. As a counterexample, consider the case where $r \ge 1$ is constant, while n tends to the infinity. In this case, the number of bins with exactly one ball tends to the constant r. Since for $r \ge 1$, the total variation distance between the Dirac measure at r and any Poisson distribution is uniformly bounded away from zero, the total variation error in the Poisson approximation cannot tend to zero.

Proof of Lemma 2.4.11. In view of Lemma 2.4.7, it suffices to prove the assertion for $m \ge 1$. First, assume that $a \le n/2$. Recalling (2.4.8), we find that

$$\begin{aligned}\lambda &\ge n\,\frac{r^m}{m!}\left(1-\frac{1}{r}\right)\left(1-\frac{2}{r}\right)\cdots\left(1-\frac{m-1}{r}\right)\frac{1}{n^m}\,e^{-a-2a/n}\\ &\ge n\,\frac{r^m}{m!}\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\cdots\left(1-\frac{m-1}{m}\right)\frac{1}{n^m}\,e^{-a-1}\\ &\ge n\left(\frac{a}{m}\right)^m e^{-a-1}.\end{aligned}$$

Recalling Remark 2.2.3, this implies $\frac{1-e^{-\lambda}}{\lambda} \leq \min\bigl\{1, \frac{e^{a+1}}{n}\bigl(\frac{m}{a}\bigr)^m\bigr\}$. On the other hand, for $a \geq n/2$, observe that $\frac{e^{a+1}}{n}\bigl(\frac{m}{a}\bigr)^m \geq \frac{m^m\, e^{a+1}}{2a^{m+1}} \geq \frac{m^m\, e^{m+2}}{2(m+1)^{m+1}} \geq \frac{e^{m+1}}{2(m+1)} \geq \frac{e}{2} \geq 1$, so that $\frac{1-e^{-\lambda}}{\lambda} \leq 1 = \min\bigl\{1, \frac{e^{a+1}}{n}\bigl(\frac{m}{a}\bigr)^m\bigr\}$. This proves the result with $B_m = e\, m^m$. □

Chapter 3

Normal approximation

3.1 Decomposable random variables

In Section 1.2, we derived that for a standard normal random variable W, the *Stein expectation*

$$\mathbb{E}\big[f'(W) - f(W)\,W\big] \tag{3.1.1}$$

vanishes for all continuously differentiable functions $f\colon \mathbb{R} \to \mathbb{R}$ of polynomial growth. Here, we shall first show that the Stein expectation can be small for random variables W featuring a certain dependence structure. In the next section, we shall show that this implies proximity to the standard normal distribution in a certain metric.

In Subsection 2.3.2, we have shown that Poisson approximation by Stein's method works well for locally dependent random variables and, more generally, random variables which can be decomposed as in (2.3.2). Here, we adjust this approach to the case of normal approximation. This adjustment is due to Barbour, Karoński and Ruciński [6].

Consider a random variable W with $\mathbb{E}\,W = 0$ and $\mathrm{var}(W) = 1$ (this, of course, means that $\mathbb{E}(W^2) < \infty$). Suppose that

$$W = \sum_{i\in\mathscr{I}} X_i\,, \tag{3.1.2}$$

where $\mathscr{I}$ is a countable set and where we assume that $\sum_{i\in\mathscr{I}} \mathbb{E}\,|X_i| < \infty$, $\sum_{i\in\mathscr{I}} \mathbb{E}\big(|X_i|\,|W|\big) < \infty$ and $\mathbb{E}\,X_i = 0$ for all $i \in \mathscr{I}$ (notice that the first condition guarantees the almost sure existence of the random sum $\sum_{i\in\mathscr{I}} X_i$). Then the variance can be expressed as

$$1 = \mathrm{var}(W) = \mathbb{E}(W^2) = \sum_{i\in\mathscr{I}} \mathbb{E}(X_i W)\,.$$

For a continuously differentiable function f with bounded derivative, this leads to the following expression of the Stein expectation:

$$\mathbb{E}\big[f'(W) - f(W)W\big] = \sum_{i\in\mathscr{I}} \mathbb{E}\Big[f'(W)\,\mathbb{E}(X_i W) - f(W)X_i\Big]$$

(notice that f is of at most linear growth, that is, there exist C_0 and C_1, such that $|f(w)| \leq C_0 + C_1|w|$ for all $w \in \mathbb{R}$). Now suppose that for each $i \in \mathscr{I}$, W can be decomposed as

$$W = W_i + R_i\,, \tag{3.1.3}$$

where W_i is independent of X_i and where $\mathbb{E}\big(|X_i|\,|R_i|\big) < \infty$. Then we have $\mathbb{E}(X_iW) = \mathbb{E}(X_iR_i)$. Next, by the fundamental theorem of calculus, we have

$$f(W) = f(W_i) + \int_{W_i}^{W} f'(t)\,\mathrm{d}t = f(W_i) + \int_0^1 f'(W_i + tR_i)R_i\,\mathrm{d}t\,.$$

Assuming in addition that $\sum_{i\in\mathscr{I}} \mathbb{E}\big(|X_i|\,|R_i|\big) < \infty$ and combining all together, we express the Stein expectation as

$$\mathbb{E}\big[f'(W) - f(W)W\big] = \sum_{i\in\mathscr{I}} \mathbb{E}\bigg[f'(W)\,\mathbb{E}(X_iR_i) - f(W_i)X_i - \int_0^1 f'(W_i + tR_i)X_iR_i\,\mathrm{d}t\bigg]\,.$$

By independence and since $\mathbb{E}\,X_i = 0$, the second term vanishes. Taking a random variable θ_1, which is uniformly distributed over $[0,1]$ and independent of all other random variables, we can rewrite the Stein expectation as

$$\mathbb{E}\big[f'(W) - f(W)W\big] = \sum_{i\in\mathscr{I}} \mathbb{E}\Big[f'(W)\,\mathbb{E}(X_iR_i) - f'(W_i + \theta_1R_i)X_iR_i\Big]\,.$$

Now suppose in addition that the random variables R_i can be expressed as sums

$$R_i = \sum_{j\in\mathscr{J}_i} X_{ij}\,,$$

where we assume that $\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i} \mathbb{E}\big(|X_i|\,|X_{ij}|\big) < \infty$. Then we can further rewrite the Stein expectation as

$$\mathbb{E}\big[f'(W) - f(W)W\big] = \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i} \mathbb{E}\Big[f'(W)\,\mathbb{E}(X_iX_{ij}) - f'(W_i + \theta_1R_i)X_iX_{ij}\Big]\,.$$

Next, suppose that for each $i \in \mathscr{I}$ and $j \in \mathscr{J}_i$, W_i can be further decomposed as

$$W_i = W_{ij} + R_{ij}\,, \tag{3.1.4}$$

where W_{ij} is independent of the pair (X_i, X_{ij}). Assuming in addition that f is twice continuously differentiable with bounded second derivative and

$$\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i} \mathbb{E}\big(|X_i|\,|X_{ij}|\big)\,\mathbb{E}\,|R_i + R_{ij}| < \infty\,,$$
$$\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i} \mathbb{E}\big(|X_i|\,|X_{ij}|\,|\theta_1R_i + R_{ij}|\big) < \infty\,,$$

we can further expand the Stein expectation as

$$\begin{aligned}\mathbb{E}\big[f'(W) - f(W)W\big] &= \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\Big[f'(W_{ij})\,\mathbb{E}(X_iX_{ij}) + f''(W_{ij}+\theta_2R_{ij}+\theta_2R_i)(R_i+R_{ij})\,\mathbb{E}(X_iX_{ij})\\ &\qquad - f'(W_{ij})X_iX_{ij} - f''(W_{ij}+\theta_2R_{ij}+\theta_1\theta_2R_i)X_iX_{ij}(\theta_1R_i+R_{ij})\Big]\,,\end{aligned}$$

where θ_2 is another random variable which is uniformly distributed over $[0,1]$ and is independent of all other random variables. Because of independence, the first and the third term cancel, so that the Stein expectation can be expressed as

$$\begin{aligned}\mathbb{E}\big[f'(W) - f(W)W\big] = \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\Big[&f''(W_{ij}+\theta_2R_{ij}+\theta_2R_i)(R_i+R_{ij})\,\mathbb{E}(X_iX_{ij})\\ &- f''(W_{ij}+\theta_2R_{ij}+\theta_1\theta_2R_i)X_iX_{ij}(\theta_1R_i+R_{ij})\Big]\,.\end{aligned}\tag{3.1.5}$$

However, the assumptions implying the preceding identity can be relaxed to a certain extent. First, the assumption that f is twice continuously differentiable with bounded second derivative can be replaced by the assumption that f is differentiable and f' is Lipschitz (see Proposition B.1.9). This is because the fundamental theorem of calculus remains true for Lipschitz test functions (and more generally for the absolutely continuous functions): see Section B.1. This allows us to bound the Stein expectation in terms of

$$M_2(f) := \operatorname{ess\,sup}|f''| = \sup_{x\neq y}\frac{|f'(x)-f'(y)|}{|x-y|}\,,$$

where f' is the classical derivative of f and f'' is an almost-everywhere derivative of f'.

In addition to the relaxed assumption on differentiability of f, one can also drop certain other assumptions. The following assertion makes it precise.

Proposition 3.1.1. *Let W be decomposed as follows:*

$$\begin{gathered}W = \sum_{i\in\mathscr{I}} X_i\,,\\ W = W_i + R_i\,,\ \textit{where } W_i \textit{ is independent of } X_i\,,\\ R_i = \sum_{j\in\mathscr{J}_i} X_{ij}\,,\\ W_i = W_{ij} + R_{ij}\,,\ \textit{where } W_{ij} \textit{ is independent of } (X_i, X_{ij})\,.\end{gathered}$$

Next, suppose that $\mathbb{E}\,X_i = 0$ for all $i\in\mathscr{I}$ and $\operatorname{var}(W) = 1$, and that

$$\begin{gathered}\sum_{i\in\mathscr{I}}\mathbb{E}\,|X_i| < \infty\,,\qquad \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big(|X_i|\,|X_{ij}|\big) < \infty\,,\qquad \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big(|X_i|\,|X_{ij}|\,|R_{ij}|\big) < \infty\,,\\ \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big(|X_i|\,|X_{ij}|\,|R_i+R_{ij}|\big) < \infty\,,\qquad \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big(|X_i|\,|X_{ij}|\big)\,\mathbb{E}\,|R_i+R_{ij}| < \infty\,.\end{gathered}\tag{3.1.6}$$

Finally, take a function $f\colon \mathbb{R} \to \mathbb{R}$ *with* $M_2(f) < \infty$, *where* $M_2(f)$ *is defined as in* (B.1.3). *Then* (3.1.5) *remains true and we can estimate*

$$\left|\mathbb{E}\big[f'(W) - f(W)W\big]\right| \leq M_2(f) \sum_{i\in\mathscr{I}} \sum_{j\in\mathscr{J}_i} \mathbb{E}\Big[|X_i|\,|X_{ij}|\Big(\mathbb{E}\,|R_i + R_{ij}| + \tfrac{1}{2}|R_{ij}| + \tfrac{1}{2}|R_i + R_{ij}|\Big)\Big]\,. \tag{3.1.7}$$

Before proving the preceding assertion, we formulate a couple of remarks. First, we no longer assume that f' is bounded. In particular, this allows us to apply Proposition 3.1.1 with the function $f(w) = w|w|$: this function is differentiable with derivative $f'(w) = 2|w|$, which is Lipschitz, but not bounded. This implies that W has finite third absolute moment. More precisely, with f as above, the following result is immediate from Proposition 3.1.1:

Corollary 3.1.2. *If* W *is as in Proposition 3.1.1, then*

$$\mathbb{E}(|W|^3) \leq 2\,\mathbb{E}\,|W| + 2\sum_{i\in\mathscr{I}} \sum_{j\in\mathscr{J}_i} \mathbb{E}\Big[|X_i|\,|X_{ij}|\Big(|R_i + R_{ij}| + \tfrac{1}{2}|R_{ij}| + \tfrac{1}{2}|R_i + R_{ij}|\Big)\Big]\,.$$

□

Remark 3.1.3. If W is standard normal, then $\mathbb{E}\,|W| = 2/\sqrt{2\pi}$ and $\mathbb{E}\big(|W|^3\big) = 4/\sqrt{2\pi}$, so that $\mathbb{E}\big(|W|^3\big) = 2\,\mathbb{E}\,|W|$.

Now we turn to the proof of Proposition 3.1.1. As the first step, we formulate and prove the the following auxiliary result:

Lemma 3.1.4. *Let* $f\colon \mathbb{R} \to \mathbb{R}$ *be absolutely continuous and let* $h\colon [0,\infty) \to [0,\infty)$ *be non-decreasing. Suppose that* $|f'| \leq h$. *Finally, let a random variable* W *be as in Proposition 3.1.1, and let* θ_1 *be uniformly distributed over* $[0,1]$ *and independent of all other random variables. If*

$$\sum_{i\in\mathscr{I}} \sum_{j\in\mathscr{J}_i} \mathbb{E}\big[h(|W_i + \theta_1 R_i|)|X_i|\,|X_{ij}|\big] < \infty\,, \tag{3.1.8}$$

then $\mathbb{E}\big|f(W)W\big| < \infty$ *and*

$$\mathbb{E}\big[f(W)W\big] = \sum_{i\in\mathscr{I}} \sum_{j\in\mathscr{J}_i} \mathbb{E}\big[f'(W_i + \theta_1 R_i)X_i X_{ij}\big]\,. \tag{3.1.9}$$

Proof. Equation (3.1.9) is derived at the beginning of this section, but under stronger conditions. A closer look reveals that the calculations are still valid if W is as in Proposition 3.1.1 and f is Lipschitz (recall that in this case, f has an almost-everywhere derivative and satisfies the fundamental theorem of calculus – see Section B.1).

Now take any absolutely continuous function f with $|f'| \leq h$. For each $n \in \mathbb{N}$, define function $\psi_n\colon [0,\infty) \to [0,1]$ as $\psi_n(t) := 1$ for $t \leq n$, $\psi_n(t) := 2 - t/n$ for $n \leq t \leq 2n$ and $\psi_n(t) := 0$ for $t \geq 2n$. Observe that ψ_n is well defined and that for each fixed n, the expression $t\,\psi_n(t)$ is bounded in t. Clearly, ψ is absolutely continuous with $\psi_n'(t) = 0$ for

$t < n$, $\psi_n'(t) = -1/n$ for $n < t < 2n$ and $\psi_n'(t) = 0$ for $t > 2n$. Observe that ψ' can be extended to $[0, \infty)$ so that the expression $t\,|\psi_n'(t)|$ is uniformly bounded in t and n.

Now let $f_n(w) := f\big(\psi_n(|w|)\,w\big)$. Applying the chain rule (for details on the validity in the context of absolutely continuous functions, see Corollary 6.5.4 of [20]), we find that f_n is absolutely continuous with

$$f_n'(w) = f'\big(\psi_n(|w|)\,w\big)\big(\psi_n'(|w|)|w| + \psi_n(|w|)\big)\,.$$

Since $t\,|\psi_n'(t)|$ is uniformly bounded in t and n and since h is non-decreasing, there exists a constant C, such that $|f_n'(w)| \le C\,h(|w|)$ for all n and w.

For each fixed n, the expression $\psi_n(|w|)\,w$ is bounded in $w \in \mathbb{R}$. Since f' is bounded on bounded sets, $|f_n(w)|$ is also bounded in $w \in \mathbb{R}$. Therefore, (3.1.9) applies with f_n in place of f.

Now observe that the functions f_n converge pointwise to f and that their derivatives f_n' converge pointwise to f' as well. Recalling (3.1.8) and applying the dominated convergence theorem, we obtain

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}\big[f_n(W)W\big] &= \lim_{n\to\infty} \sum_{i\in\mathscr{I}} \sum_{j\in\mathscr{J}_i} \mathbb{E}\big[f_n'(W_i + \theta_1 R_i)X_i X_{ij}\big] \\ &= \sum_{i\in\mathscr{I}} \sum_{j\in\mathscr{J}_i} \mathbb{E}\big[f'(W_i + \theta_1 R_i)X_i X_{ij}\big]\,.\end{aligned} \tag{3.1.10}$$

Now consider the function $\tilde{f}\colon \mathbb{R} \to \mathbb{R}$ defined by $\tilde{f}(w) := \int_0^w h(|t|)\,\mathrm{d}t$. Observe that $\tilde{f}$ is absolutely continuous with $\tilde{f}'(w) = h(|w|)$ and that $\tilde{f}(w)w \ge 0$ for all $w \in \mathbb{R}$. Letting $\tilde{f}_n(w) := \tilde{f}\big(\psi_n(|w|)\,w\big)$, we also have $\tilde{f}_n(w)w \ge 0$ for all $w \in \mathbb{R}$. In addition, (3.1.10) applies with $\tilde{f}_n$ and $\tilde{f}$ in place of f_n and f. Combining with Fatou's lemma, we find that

$$\begin{aligned}\mathbb{E}\big[\tilde{f}(W)W\big] &\le \lim_{n\to\infty} \mathbb{E}\big[\tilde{f}_n(W)W\big] \\ &= \sum_{i\in\mathscr{I}} \sum_{j\in\mathscr{J}_i} \mathbb{E}\big[\tilde{f}'(W_i + \theta_1 R_i)X_i X_{ij}\big] \\ &\le \sum_{i\in\mathscr{I}} \sum_{j\in\mathscr{J}_i} \mathbb{E}\big[h(|W_i + \theta_1 R_i|)|X_i|\,|X_{ij}|\big] \\ &< \infty\,.\end{aligned} \tag{3.1.11}$$

Now estimate

$$|f_n(w)| = \big|f\big(\psi_n(|w|)\,w\big)\big| \le |f(0)| + \int_0^{\psi_n(|w|)\,|w|} h(t)\,\mathrm{d}t \le |f(0)| + \int_0^{|w|} h(t)\,\mathrm{d}t\,,$$

$$|f_n(w)w| \le |f(0)|\,|w| + |w| \int_0^{|w|} h(t)\,\mathrm{d}t \le |F(0)|\,|w| + \tilde{f}(w)w\,.$$

Recalling (3.1.11), it follows that the sequence of random variables $f_n(W)W$ is dominated by a non-negative random variable with finite expectation. Applying the dominated convergence theorem and combining with (3.1.10), finiteness of $\mathbb{E}\big|f(W)W\big|$ along with (3.1.9) follows. □

Proof of Proposition 3.1.1. We shall go from (3.1.5) backwards. Define

$$\begin{aligned}\rho := \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i} \mathbb{E}\Big[& f''(W_{ij}+\theta_2 R_{ij}+\theta_2 R_i)(R_i+R_{ij})\,\mathbb{E}(X_iX_{ij}) \\ & - f''(W_{ij}+\theta_1\theta_2 R_i+\theta_2 R_{ij})X_iX_{ij}(\theta_1 R_i+R_{ij})\Big]\,.\end{aligned} \tag{3.1.12}$$

Noting also that

$$\theta_1 R_i + R_{ij} = (1-\theta_1)R_{ij} + \theta_1(R_i+R_{ij})\,, \tag{3.1.13}$$

and applying (3.1.6), we find that the right hand side of (3.1.12) exists, along with the bound

$$|\rho| \le M\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\Big[|X_i|\,|X_{ij}|\Big(\mathbb{E}\,|R_i+R_{ij}| + \tfrac{1}{2}|R_{ij}| + \tfrac{1}{2}|R_i+R_{ij}|\Big)\Big]\,.$$

It remains to show that $\mathbb{E}\big|f'(W)-f(W)W\big|<\infty$ and $\rho = \mathbb{E}\big[f'(W)-f(W)W\big]$. Observe first that

$$\begin{aligned}\mathbb{E}\big(|X_i|\,|X_{ij}|\,|W_{ij}|\big) &= \mathbb{E}\big(|X_i|\,|X_{ij}|\big)\,\mathbb{E}\,|W_{ij}| = \\ &\le \mathbb{E}\big(|X_i|\,|X_{ij}|\big)\,\mathbb{E}\,|W| + \mathbb{E}\big(|X_i|\,|X_{ij}|\big)\big]\,\mathbb{E}\,|R_i+R_{ij}|\,.\end{aligned}$$

Applying (3.1.6), we find that

$$\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big(|X_i|\,|X_{ij}|\,|W_{ij}|\big) = \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big(|X_i|\,|X_{ij}|\big)\,\mathbb{E}\,|W_{ij}| < \infty\,.$$

Noting also that

$$|f'(w)| \le |f'(0)| + M|w|\,, \tag{3.1.14}$$

we find that

$$\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big(|f'(W_{ij})|\,|X_i|\,|X_{ij}|\big) < \infty$$

and

$$\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big[|f'(W_{ij})|\big]\big|\mathbb{E}(X_iX_{ij})\big| < \infty\,.$$

Moreover,

$$\begin{aligned}\rho = \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\Big[& f'(W_{ij})\,\mathbb{E}(X_iX_{ij}) + f''(W_{ij}+\theta_2 R_{ij}+\theta_2 R_i)(R_i+R_{ij})\,\mathbb{E}(X_iX_{ij}) \\ & - f'(W_{ij})X_iX_{ij} - f''(W_i+\theta_2 R_{ij}+\theta_1\theta_2 R_i)X_iX_{ij}(\theta_1 R_i+R_{ij})\Big]\end{aligned}$$

(and all expectations exist and the sum converges). Next, since the fundamental theorem of calculus holds for f' and f'', we have

$$\begin{aligned}&\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big[|f'(W)|\,|\,\mathbb{E}(X_iX_{ij})|\big] \\ &\quad\le \sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\Big(\mathbb{E}\big[|f'(W_{ij})|\,|\,\mathbb{E}(X_iX_{ij})|\big] + M\,\mathbb{E}\big(|X_i|\,|X_{ij}|\big)\,\mathbb{E}\,|R_i+R_{ij}|\big]\Big) < \infty\,.\end{aligned}$$

Similarly, recalling (3.1.13), we have

$$\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big[|f'(W_i+\theta_1R_i)|\,|X_i|\,|X_{ij}|\big]$$
$$\leq\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\Big(\mathbb{E}\big[|f'(W_{ij})|\,|X_i|\,|X_{ij}|\big]+\tfrac{1}{2}M\,\mathbb{E}\big[|X_i|\,|X_{ij}|\big(|R_{ij}|+|R_i+R_{ij}|\big)\big]\Big)<\infty\,.$$

In particular, the choice $f(w)=\frac{1}{2}w^2$ gives

$$\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big(|W_i+\theta_1R_i|\,|X_i|\,|X_{ij}|\big)<\infty\,. \tag{3.1.15}$$

Moreover,

$$\rho=\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\Big[f'(W)\,\mathbb{E}(X_iX_{ij})-f'(W_i+\theta_1R_i)X_iX_{ij}\Big]\,.$$

Recalling (3.1.14) and applying Lemma 3.1.4 with $h(t)=f(0)+Mt$, making use of (3.1.15), we obtain that $\mathbb{E}\big|f(W)W\big|<\infty$ and $\mathbb{E}\big[f(W)W\big]=\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big[f'(W_i+\theta_1R_i)X_iX_{ij}\big]$. Finally, applying Lemma 3.1.4 with $f(w)=w$, we find that $\mathbb{E}(W^2)=\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}(X_iX_{ij})$. Together with (3.1.14), this completes the proof. □

3.2 Solution to the Stein equation

As indicated in Section 1.2, the proximity to the standard normal distribution can be assessed as follows: for a test function h, find a function f solving the Stein equation

$$f'(w)-f(w)\,w=h(w)-\mathcal{N}h \tag{3.2.1}$$

where

$$\mathcal{N}h:=\big\langle h\,,\,\mathcal{N}(0,1)\big\rangle=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}h(x)\,e^{-x^2/2}\,\mathrm{d}x\,.$$

Taking a random variable W, we then have

$$\mathbb{E}\big[f'(W)-f(W)\,W\big]=\mathbb{E}\big[h(W)\big]-\mathcal{N}h\,. \tag{3.2.2}$$

Estimating the left hand side for sufficiently many test functions h, we are able to bound the error in the normal approximation with respect to a suitable metric. In particular, if W is as in Proposition 3.1.1 and if there exists a class $\mathscr{H}$ of test functions, such that for each $h\in\mathscr{H}$, there exists f which solves (3.2.1) and satisfies $M_2(f)\leq 1$, then we have a bound

$$d_{\mathscr{H}}\big(\mathscr{L}(W),\,\mathcal{N}(0,1)\big)\leq\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\Big[|X_i|\,|X_{ij}|\Big(\mathbb{E}\,|R_i+R_{ij}|+\tfrac{1}{2}|R_{ij}|+\tfrac{1}{2}|R_i+R_{ij}|\Big)\Big]\,,$$

where $d_{\mathscr{H}}$ is defined as in (A.1.1).

Equation (3.2.1) is an ordinary differential equation of the first order. Such differential equations can be solved in two steps: first, we solve the homogeneous part, then we perform the variation of constant. A simple calculation shows that the solution to the homogeneous part

$$f_{\mathrm{H}}'(w) - f_{\mathrm{H}}(w)\, w = 0 \tag{3.2.3}$$

is $f_{\mathrm{H}}(w) = C\, e^{w^2/2}$. The solution to the original equation (3.2.1) can be seeked as $f(w) = k(w)\, e^{w^2/2}$, where k is now a function. Another simple calculation shows that k must be the indefinite integral

$$k(w) = \int \big(h(x) - \mathcal{N}h\big)\, e^{-x^2/2}\, \mathrm{d}x\,.$$

However, unless $\lim_{w\to\pm\infty} k(w) = 0$, f grows very rapidly, so that there is no hope for f' to be Lipschitz. Since $\int_{-\infty}^{\infty} \big(h(x) - \mathcal{N}h\big)\, \mathrm{d}x = 0$, $\lim_{w\to-\infty} k(w) = 0$ implies $\lim_{w\to\infty} k(w) = 0$ and the 'tame' solution to (3.2.1) can be expressed as

$$f(w) = e^{w^2/2} \int_{-\infty}^{w} \big(h(x) - \mathcal{N}h\big)\, e^{-x^2/2}\, \mathrm{d}x = e^{w^2/2} \int_{w}^{\infty} \big(\mathcal{N}h - h(x)\big)\, e^{-x^2/2}\, \mathrm{d}x\,. \tag{3.2.4}$$

We summarize the preceding calculations into the following statement:

Proposition 3.2.1. *For any measurable function $h\colon \mathbb{R} \to \mathbb{R}$ with $\mathcal{N}|h| < \infty$, the function f defined by (3.2.4) is an almost-everywhere solution to the Stein equation (3.2.1), i. e., f is absolutely continuous and the function $w \mapsto f(w)\, w + h(w) - \mathcal{N}h$ is an almost-everywhere derivative of f. If h is continuous, f is a classical solution, i. e., continuously differentiable and (3.2.1) holds for all $w \in \mathbb{R}$.*

Denoting the standard normal density by

$$\phi(z) := \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\,,$$

we can also write

$$f(w) = \frac{1}{\phi(w)} \int_{-\infty}^{w} \big(h(x) - \mathcal{N}h\big)\, \phi(x)\, \mathrm{d}x = \frac{1}{\phi(w)} \int_{w}^{\infty} \big(\mathcal{N}h - h(x)\big)\, \phi(x)\, \mathrm{d}x\,.$$

Of course, we can also take affine combinations of the two forms, that is, for each $a \in \mathbb{R}$, we have

$$\begin{aligned} f(w) &= \frac{1}{\phi(w)} \left((1-a) \int_{-\infty}^{w} \big(h(x) - \mathcal{N}h\big)\, \phi(x)\, \mathrm{d}x + a \int_{w}^{\infty} \big(\mathcal{N}h - h(x)\big)\, \phi(x)\, \mathrm{d}x \right) \\ &= \frac{1}{\phi(w)} \left((1-a) \int_{-\infty}^{w} h(x)\, \phi(x)\, \mathrm{d}x - a \int_{w}^{\infty} h(x)\, \phi(x)\, \mathrm{d}x \right) \\ &\qquad + \big[a(1-\Phi(w)) - (1-a)\Phi(w)\big]\mathcal{N}h\,, \end{aligned} \tag{3.2.5}$$

where

$$\Phi(w) := \int_{-\infty}^{\infty} \phi(x)\, \mathrm{d}x$$

is the standard normal cumulative distribution function. Choosing $a = \Phi(w)$, the second term in the final expression of (3.2.5) vanishes and we obtain the following form:

$$f(w) = \frac{1-\Phi(w)}{\phi(w)} \int_{-\infty}^{w} h(x)\,\phi(x)\,\mathrm{d}x - \frac{\Phi(w)}{\phi(w)} \int_{w}^{\infty} h(x)\,\phi(x)\,\mathrm{d}x\,.$$

Noting that $1-\Phi(w) = \Phi(-w)$ and introducing the *Mills ratio*:

$$\psi(w) := \frac{\Phi(w)}{\phi(w)}$$

we can rewrite the solution as

$$f(w) = \psi(-w) \int_{-\infty}^{w} h(x)\,\phi(x)\,\mathrm{d}x - \psi(w) \int_{w}^{\infty} h(x)\,\phi(x)\,\mathrm{d}x\,.$$

In view of Proposition 3.1.1, it is beneficial to study the behaviour of the second derivative, including the general case where f' is absolutely continuous. Differentiating the preceding formula, we obtain

$$\begin{aligned} f'(w) &= -\psi'(-w) \int_{-\infty}^{w} h(x)\,\phi(x)\,\mathrm{d}x + \psi(-w)\,h(w)\,\phi(w) \\ &\quad - \psi'(w) \int_{w}^{\infty} h(x)\,\phi(x)\,\mathrm{d}x + \psi(w)\,h(w)\,\phi(w) \\ &= h(w) - \psi'(-w) \int_{-\infty}^{w} h(x)\,\phi(x)\,\mathrm{d}x - \psi'(w) \int_{w}^{\infty} h(x)\,\phi(x)\,\mathrm{d}x\,, \end{aligned} \tag{3.2.6}$$

where the last equality follows from the identity $\Phi(w) + \Phi(-w) = 1$ (this is also Proposition C.2.3 for the case $r = 0$).

Remark 3.2.2. Similarly as in Proposition 3.2.1, Formula 3.2.6 can be interpreted in two ways: if h is continuous, f' defined as in (3.2.6) is the classical derivative of f, whereas in the general case, f is absolutely continuous and f' is an almost-everywhere derivative of f.

Now assume that h is absolutely continuous (it must be if so is f') and that it is of polynomial growth. In this case, we can apply the integration by parts formula (see Theorem 6.4.6 of Heil [20]), where we differentiate h and integrate ϕ. However, in the first integral of (3.2.6), we integrate $\phi(x)$ to $\Phi(x)$, while in the second one, we integrate $\phi(x)$ to $-\Phi(-x)$. Noting that the polynomial growth of h along with Corollary C.2.2 implies $\lim_{w\to-\infty} h(w)\,\Phi(w) = 0$ and $\lim_{w\to\infty} h(w)\,\Phi(-w) = 0$, we obtain

$$\begin{aligned} f'(w) &= h(w) - \psi'(-w)\,h(w)\,\Phi(w) + \psi'(-w) \int_{-\infty}^{w} h'(x)\,\Phi(x)\,\mathrm{d}x \\ &\quad - \psi'(w)\,h(w)\,\Phi(-w) - \psi'(w) \int_{w}^{\infty} h'(x)\,\Phi(-x)\,\mathrm{d}x \\ &= \psi'(-w) \int_{-\infty}^{w} h'(x)\,\Phi(x)\,\mathrm{d}x - \psi'(w) \int_{w}^{\infty} h'(x)\,\Phi(-x)\,\mathrm{d}x\,, \end{aligned} \tag{3.2.7}$$

where the last inequality is due to Proposition C.2.3 for $r = 1$. Differentiating (3.2.7), we find that

$$\begin{aligned} f''(w) &= -\psi''(-w) \int_{-\infty}^{w} h'(x)\,\Phi(x)\,\mathrm{d}x + \psi'(-w)\,h'(w)\,\Phi(w)\,\mathrm{d}w \\ &\quad - \psi''(w) \int_{w}^{\infty} h'(x)\,\Phi(-x)\,\mathrm{d}x + \psi'(w)\,h'(w)\,\Phi(-w) \\ &= h'(w) - \psi''(-w) \int_{-\infty}^{w} h'(x)\,\Phi(x)\,\mathrm{d}x - \psi''(w) \int_{w}^{\infty} h'(x)\,\Phi(-x)\,\mathrm{d}x\,, \end{aligned} \tag{3.2.8}$$

where the last inequality again follows from Proposition C.2.3 for $r = 1$. Again, if h' is continuous, f'' is the classical derivative of f', whereas in the general case, f' is absolutely continuous and f'' is an almost-everywhere derivative of f'.

The preceding formula allows us to bound $M_2(f)$ in terms of $M_1(f)$. For random variables decomposed according to Barbour, Karoński and Ruciński, this allows us to bound the error in the normal approximation in the Wasserstein metric.

Theorem 3.2.3. *For any function $h\colon \mathbb{R} \to \mathbb{R}$ with $M_1(h) < \infty$, the function f defined by (3.2.4) is a classical solution to the Stein equation (3.2.1), which satisfies $M_2(f) \leq M_1(h)$.*

Corollary 3.2.4. *For a random variable W decomposed as in Proposition 3.1.1, we have*

$$d_{\mathrm{W}}\big(\mathscr{L}(W), \mathcal{N}(0,1)\big) \leq \sum_{i\in\mathscr{I}} \sum_{j\in\mathscr{J}_i} \mathbb{E}\Big[|X_i|\,|X_{ij}|\Big(2\,\mathbb{E}\,|R_i + R_{ij}| + |R_{ij}| + |R_i + R_{ij}|\Big)\Big]\,. \tag{3.2.9}$$

Proof of Theorem 3.2.3. Since $M_1(h) < \infty$, h is absolutely continuous and of linear growth. Consequently, $\mathcal{N}|h| < \infty$ and Proposition 3.2.1 applies. Moreover, since h is absolutely continuous and of linear growth, the derivation of (3.2.8) is valid. By Proposition C.1.5 and Proposition C.2.3 for $r = 2$, we can estimate

$$|f''(w)| \leq M_1(h) \left[1 + \psi''(-w) \int_{-\infty}^{w} \Phi(x)\,\mathrm{d}x + \psi''(w) \int_{w}^{\infty} \Phi(-x)\,\mathrm{d}x\right] = 2\,M_1(f)\,.$$

Taking the supremum over w, the proof is complete. □

Example 3.2.5. Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed random variables with $\mathbb{E}\,\xi_1 = 0$, $\mathrm{var}(\xi_1) = 1$ and $\mathbb{E}\,|\xi_1|^3 < \infty$. Then the rescaled sum

$$W^{(n)} := \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{\sqrt{n}}$$

satisfies $\mathbb{E}\,W^{(n)} = 0$ and $\mathrm{var}\big(W^{(n)}\big) = 1$, and can be trivially decomposed by setting $\mathscr{I}^{(n)} := \{1, 2, \ldots, n\}$, $X_i^{(n)} := \xi_i/\sqrt{n}$, $R_i^{(n)} := X_i^{(n)}$, $\mathscr{J}_i^{(n)} := \{0\}$, $X_{i0}^{(n)} := X_i^{(n)}$ and $R_{i0}^{(n)} := 0$. Corollary 3.2.4 yields

$$d_{\mathrm{W}}\big(\mathscr{L}\big(W^{(n)}\big),\, \mathcal{N}(0,1)\big) \leq \sum_{i=1}^{n} \Big(2\,\mathbb{E}\big(X_i^{(n)}\big)^2\,\mathbb{E}\big|X_i^{(n)}\big| + \mathbb{E}\big|X_i^{(n)}\big|^3\Big)\,.$$

By Jensen's inequality, we have $\mathbb{E}\big|X_i^{(n)}\big| \le \Big(\mathbb{E}\big|X_i^{(n)}\big|^3\Big)^{1/3}$ and $\mathbb{E}\big(X_i^{(n)}\big)^2 \le \Big(\mathbb{E}\big|X_i^{(n)}\big|^3\Big)^{2/3}$, leading to

$$d_{\mathrm{W}}\big(\mathscr{L}\big(W^{(n)}\big),\, \mathcal{N}(0,1)\big) \le 3\sum_{i=1}^{n} \mathbb{E}\big|X_i^{(n)}\big|^3 = \frac{3\,\mathbb{E}\,|\xi_1|^3}{\sqrt{n}}\,.$$

This is the typical rate of convergence in the central limit theorem and can in general not be improved. As an example, take $\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = -1) = 1/2$. For even n, $W^{(n)}$ takes values in the set $\big\{2k/\sqrt{n}\,;\, k \in \mathbb{Z}\big\}$. Now define functions $f_n\colon \mathbb{R} \to \mathbb{R}$ by $f_n(w) := \big|w - \frac{2k}{\sqrt{n}}\big|$ for $\frac{2k-1}{\sqrt{n}} \le w \le \frac{2k+1}{\sqrt{n}}$, where $k \in \mathbb{Z}$. Then we have $\mathbb{E}\big[f_n\big(W^{(n)}\big)\big] = 0$. Observe that

$$\begin{aligned}
\mathcal{N}f_n &= \int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,\mathrm{d}x \\
&= \frac{1}{2}\int_{-\infty}^{\infty} \left[f_n(x)\,\phi(x)\,\mathrm{d}x + f_n\left(x+\frac{1}{\sqrt{n}}\right)\phi\left(x+\frac{1}{\sqrt{n}}\right)\right]\mathrm{d}x \\
&= \frac{1}{2}\int_{-\infty}^{\infty} \left[f_n(x)\,\phi(x)\,\mathrm{d}x + f_n\left(x+\frac{1}{\sqrt{n}}\right)\phi\left(x+\frac{1}{\sqrt{n}}\right)\right]\mathrm{d}x \\
&= \frac{1}{2}\int_{-\infty}^{\infty} \left[f_n(x) + f_n\left(x+\frac{1}{\sqrt{n}}\right)\right]\phi(x)\,\mathrm{d}x \\
&\qquad\qquad + \frac{1}{2}\int_{-\infty}^{\infty} f_n\left(x+\frac{1}{\sqrt{n}}\right)\left[\phi\left(x+\frac{1}{\sqrt{n}}\right) - \phi(x)\right]\mathrm{d}x \\
&= \frac{1}{2\sqrt{n}} + \frac{1}{2\sqrt{n}}\int_{-\infty}^{\infty} f_n\left(x+\frac{1}{\sqrt{n}}\right)\int_0^1 \phi'\left(x+\frac{t}{\sqrt{n}}\right)\mathrm{d}t\mathrm{d}x\,.
\end{aligned}$$

Noting that $0 \le f_n(x) \le 1/\sqrt{x}$ for all x and that the function ϕ' is bounded, we find that $\lim_{n\to\infty} \mathcal{N}f_n\,\sqrt{n} = 1/2$. Therefore, for each $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$, such that $d_{\mathrm{W}}\big(\mathscr{L}(W_n),\, \mathcal{N}(0,1)\big) \ge \frac{1-\varepsilon}{2\sqrt{n}}$. This proves that the rate of $1/\sqrt{n}$ cannot be improved.

Remark 3.2.6. The bound in Theorem 3.2.3 is sharp: consider functions

$$h_n(w) := \begin{cases} w + \frac{2}{n} & ;\, w \le -\frac{1}{n} \\ -w & ;\, -\frac{1}{n} \le w \le \frac{1}{n} \\ w - \frac{2}{n} & ;\, w \ge \frac{1}{n}\,. \end{cases}$$

Clearly, $M_1(h_n) = 1$. Next, observe that the underlying functions f_n'' are continuous at the origin and we have

$$f_n''(0) = -1 - \psi''(0)\int_{-\infty}^{0} h_n'(x)\,\Phi(x)\,\mathrm{d}x - \psi''(0)\int_0^{\infty} h_n'(x)\,\Phi(-x)\,\mathrm{d}x\,.$$

Therefore,

$$M_2(f_n) \ge \left|1 + \psi''(0)\int_{-\infty}^{0} h_n'(x)\,\Phi(x)\,\mathrm{d}x + \psi''(0)\int_0^{\infty} h_n'(x)\,\Phi(-x)\,\mathrm{d}x\right|\,.$$

The functions h_n' are uniformly bounded and converge pointwise to the constant 1 (except at 0, where they converge to -1). By the dominated convergence theorem, the right hand

side converges to

$$\left|1+\psi''(0)\int_{-\infty}^{0}\Phi(x)\,\mathrm{d}x+\psi''(0)\int_{0}^{\infty}\Phi(-x)\,\mathrm{d}x\right|=2$$

by Proposition C.2.3 for $r=2$. Therefore, for each $\varepsilon>0$, there exists n such that $M_2(f_n)>2-\varepsilon$.

3.3 Applications

3.3.1 Local dependence and U-statistics

In Subsection 2.3.2, we already considered locally dependent random variables. The concept of local dependence was expressed in terms of dependence neighbourhoods. Here, we shall take a stronger concept expressed in terms of the *dependence graph*.

For two vertices i and j of an undirected graph Γ, we shall denote $i\sim j$ if they are equal or adjacent, and $i\not\sim j$ otherwise. For a vertex i and a set of vertices $\mathscr{J}$, we shall denote $i\sim\mathscr{J}$ if either $i\in$ or there is an edge with one endpoint equal to i and the other in $\mathscr{J}$, and $i\not\sim\mathscr{J}$ otherwise. Finally, for sets of vertices $\mathscr{I}$ and $\mathscr{J}$, we shall denote $\mathscr{I}\sim\mathscr{J}$ if either $\mathscr{I}\cap\mathscr{J}\neq\emptyset$ or there is an edge with one endpoint in $\mathscr{I}$ and the other in $\mathscr{J}$, and $\mathscr{I}\not\sim\mathscr{J}$ otherwise.

Definition 3.3.1. Let $(X_i)_{i\in\mathscr{I}}$ be a family of random variables and let Γ be an undirected graph with vertex set $\mathscr{I}$. The dependence structure of the family $(X_i)_{i\in\mathscr{I}}$ fits Γ if for any sets $\mathscr{J},\mathscr{K}\subseteq\mathscr{I}$ with $\mathscr{J}\not\sim\mathscr{K}$, the subfamilies $(X_j)_{j\in\mathscr{J}}$ and $(X_k)_{k\in\mathscr{K}}$ are independent. We shall call Γ a *dependence graph* for the family $(X_i)_{i\in\mathscr{I}}$.

As in Section 3.1, consider a sum $W=\sum_{i\in\mathscr{I}}X_i$, such that $\sum_{i\in\mathscr{I}}\mathbb{E}\,|X_i|<\infty$, $\mathbb{E}\,X_i=0$ for all $i\in\mathscr{I}$ and $\mathrm{var}(W)=1$; in addition, suppose that $\sum_{i\in\mathscr{I}}\mathbb{E}\,|X_i|^3<\infty$. Next, let Γ be a dependence graph for the family $(X_i)_{i\in\mathscr{I}}$. Take $D<\infty$ and suppose that for each $i\in\mathscr{I}$, there are no more than D vertices j with $i\sim j$. In other words, the degrees of all vertices are strictly less than D. We shall show that under these conditions, W can be reasonably decomposed as in Proposition 3.1.1. To achieve this, first set

$$\mathscr{J}_i:=\{j\in\mathscr{I}\,;\,i\not\sim j\}\,,\qquad X_{ij}:=X_j\,,\qquad R_i:=\sum_{j;i\sim j}X_j\,,\qquad W_i:=\sum_{j;i\not\sim j}X_j\,.$$

Since the dependence structure of the family $(X_i)_{i\in\mathscr{I}}$ fits Γ, W_i is independent of X_i. Next, set

$$R_{ij}:=\sum_{k;i\not\sim k,j\sim k}X_k\,,\qquad W_{ij}:=\sum_{k;k\not\sim\{i,j\}}X_k$$

and again observe that W_{ij} is independent of the pair (X_i,X_j). Now we bound the quantities appearing in (3.1.6) and (3.1.7). First, by the inequality between the arithmetic and the geometric mean, we have

$$\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}|X_i|\,|X_{ij}|=\sum_{(i,j);i\sim j}|X_i|\,|X_j|\leq\frac{1}{2}\sum_{(i,j);i\sim j}\left(X_i^2+X_j^2\right).$$

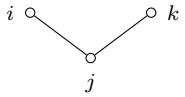
Since the vertex degrees are bounded from above by D, we have $\big|\{j\ ;\ i \sim j\}\big| \leq D$ and $\big|\{j\ ;\ i \sim j\}\big| \leq D$, leading to

$$\sum_{i \in \mathscr{I}} \sum_{j \in \mathscr{J}_i} |X_i|\,|X_{ij}| \leq D \sum_{i \in \mathscr{I}} \mathbb{E}(X_i^2) \leq \frac{D}{2} \sum_{i \in \mathscr{I}} \Big(\mathbb{E}\,|X_i| + \mathbb{E}\,|X_i|^3 \Big) < \infty\,,$$

applying again the inequality between the arithmetic and the geometric mean in the second inequality. Next, observe that

$$\sum_{i \in \mathscr{I}} \sum_{j \in \mathscr{J}_i} |X_i|\,|X_{ij}|\,|R_{ij}| = \sum_{(i,j,k) \in \mathscr{T}_1} |X_i|\,|X_j|\,|X_k| \leq \frac{1}{3} \sum_{(i,j,k) \in \mathscr{T}_1} \big(|X_i|^3 + |X_j|^3 + |X_k|^3 \big)\,,$$

where $\mathscr{T}_1 := \big\{(i,j,k)\ ;\ i \sim j, j \sim k, i \not\sim k\big\}\big\}$ as illustrated below:



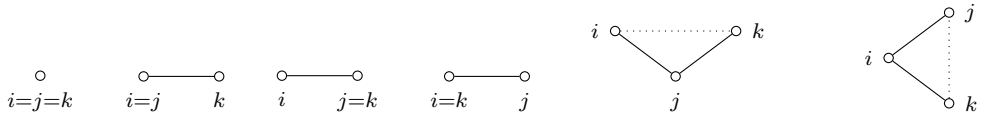
The elements i, j and k must be distinct. Therefore, $\big|\{(j,k)\ ;\ (i,j,k) \in \mathscr{T}_1\}\big| \leq (D-1)(D-2)$ for all i, $\big|\{(i,k)\ ;\ (i,j,k) \in \mathscr{T}_1\}\big| \leq (D-1)(D-2)$ for all j and $\big|\{(i,j)\ ;\ (i,j,k) \in \mathscr{T}_1\}\big| \leq (D-1)(D-2)$ for all k, leading to

$$\sum_{i \in \mathscr{I}} \sum_{j \in \mathscr{J}_i} \mathbb{E}\big(|X_i|\,|X_{ij}|\,|R_{ij}| \big) \leq (D-1)(D-2) \sum_{i \in \mathscr{I}} \mathbb{E}\,|X_i|^3\,.$$

Next, observe that

$$\sum_{i \in \mathscr{I}} \sum_{j \in \mathscr{J}_i} |X_i|\,|X_{ij}|\,|R_i + R_{ij}| = \sum_{(i,j,k) \in \mathscr{T}_2} |X_i|\,|X_j|\,|X_k| \leq \frac{1}{3} \sum_{(i,j,k) \in \mathscr{T}_2} \big(|X_i|^3 + |X_j|^3 + |X_k|^3 \big)\,,$$

where $\mathscr{T}_2 := \big\{(i,j,k)\ ;\ i \sim j, i \sim \{j,k\}\big\}$. Elements (i,j,k) of the set $\mathscr{T}_2$ can be divided into the cases illustrated below, along with upper bounds on the number of pairs (j,k) with $(i,j,k) \in \mathscr{T}_2$ for fixed i, which are also upper bounds on the number of pairs (i,j) with $(i,j,k) \in \mathscr{T}_2$ for fixed j and upper bounds on the number of pairs (i,j) with $(i,j,k) \in \mathscr{T}_2$ for fixed k:



1	$D-1$	$D-1$	$D-1$	$(D-1)(D-2)$	$(D-1)(D-2)$

Therefore, $\big|\{(j,k)\ ;\ (i,j,k) \in \mathscr{T}_2\}\big| \leq 2D^2 - 3D + 2$ for all i, $\big|\{(i,k)\ ;\ (i,j,k) \in \mathscr{T}_2\}\big| \leq 2D^2 - 3D + 2$ for all j and $\big|\{(i,j)\ ;\ (i,j,k) \in \mathscr{T}_2\}\big| \leq 2D^2 - 3D + 2$ for all k, leading to

$$\sum_{i \in \mathscr{I}} \sum_{j \in \mathscr{J}_i} \mathbb{E}\big(|X_i|\,|X_{ij}|\,|R_i + R_{ij}| \big) \leq (2D^2 - 3D + 2) \sum_{i \in \mathscr{I}} \mathbb{E}\,|X_i|^3\,.$$

Writing $\mathbb{E}\big(|X_i|\,|X_{ij}|\big)\,\mathbb{E}\,|R_i + R_{ij}| = \mathbb{E}\big(|X_i|\,|X_{ij}|\,|R'_i + R'_{ij}|\big)$, where the pair (R'_i, R'_{ij}) is an independent copy of the pair (R_i, R_{ij}), we can similarly estimate

$$\sum_{i\in\mathscr{I}}\sum_{j\in\mathscr{J}_i}\mathbb{E}\big(|X_i|\,|X_{ij}|\big)\,\mathbb{E}\,|R_i + R_{ij}| \le (2D^2 - 3D + 2)\sum_{i\in\mathscr{I}}\mathbb{E}\,|X_i|^3\,.$$

Collecting all together and applying Corollary 3.2.4, we obtain the following result:

Theorem 3.3.2. *Let $(X_i)_{i\in\mathscr{I}}$ be a family of random variables with dependence structure which fits a graph Γ. Suppose that for each $i \in \mathscr{I}$, there are no more than $D < \infty$ vertices j with $i \sim j$. Assume that $\sum_{i\in\mathscr{I}}\mathbb{E}\,|X_i| < \infty$, $\sum_{i\in\mathscr{I}}\mathbb{E}\,|X_i|^3 < \infty$ and $\mathbb{E}\,X_i = 0$ for all $i \in \mathscr{I}$. Let $W = \sum_{i\in\mathscr{I}} X_i$ and suppose that $\operatorname{var}(W) = 1$. Then we have*

$$d_{\mathrm{W}}\big(\mathscr{L}(W),\,\mathcal{N}(0,1)\big) \le (7D^2 - 12D + 8)\sum_{i\in\mathscr{I}}\mathbb{E}\,|X_i|^3\,.$$

□

Example 3.3.3. If the summands X_i are independent, we can set $i \sim j$ if and only if $i = j$, leading to $D = 1$ and the bound

$$d_{\mathrm{W}}\big(\mathscr{L}(W),\,\mathcal{N}(0,1)\big) \le 3\sum_{i\in\mathscr{I}}\mathbb{E}\,|X_i|^3\,,$$

which is the same as the bound in Example 3.2.5.

Example 3.3.4. Consider *U-statistics*: let $\xi_1, \xi_2, \ldots$ be independent and identically distributed random variables taking values in a measurable space $(S, \mathscr{S})$. Let $F\colon S\times S \to \mathbb{R}$ be a symmetric product measurable function. Suppose that $\mathbb{E}\big|F(\xi_1,\xi_2)\big|^3 < \infty$ and $\mathbb{E}\,F(\xi_1,\xi_2) = 0$. Consider the sum

$$U_n := \sum_{1\le i<j\le n} F(\xi_i,\xi_j)\,.$$

Now compute the variance:

$$\sigma_n^2 := \operatorname{var}(U_n) = \sum_{1\le i<j\le n}\sum_{1\le j<k\le n}\operatorname{cov}\big(F(\xi_i,\xi_j),\ F(\xi_k,\xi_l)\big)\,.$$

Noting that

$$\operatorname{cov}\big(F(\xi_i,\xi_j),\ F(\xi_k,\xi_l)\big) = \begin{cases} \operatorname{var}\big(F(\xi_1,\xi_2)\big) & ;\ i=j,\ k=l \\ \operatorname{cov}\big(F(\xi_1,\xi_2),\ F(\xi_1,\xi_3)\big) & ;\ |\{i,j\}\cap\{k,l\}| = 1 \\ 0 & ;\ \text{otherwise} \end{cases}$$

and letting $V_1 := \operatorname{var}\big(F(\xi_1,\xi_2)\big)$ and $V_2 := \operatorname{cov}\big(F(\xi_1,\xi_2)\,F(\xi_1,\xi_3)\big)$, we obtain

$$\sigma_n^2 = \frac{n(n-1)}{2}\,V_1 + n(n-1)(n-2)\,V_2\,.$$

Assume that F is not almost everywhere zero (with respect to the joint distribution of (ξ_1, ξ_2), so that $V_1 > 0$. The covariance V_2 can be decomposed as

$$V_2 = \mathbb{E}\Big[\operatorname{cov}\big(F(\xi_1,\xi_2),\, F(\xi_1,\xi_3) \;\big|\; \xi_1\big)\Big] + \operatorname{cov}\Big[\mathbb{E}\big(F(\xi_1,\xi_2) \;\big|\; \xi_1\big),\, \mathbb{E}\big(F(\xi_1,\xi_3) \;\big|\; \xi_1\big)\Big]\,.$$

Since ξ_2 and ξ_3 are conditionally independent give ξ_1, the first term vanishes. Next, $\mathbb{E}\big(F(\xi_1,\xi_2) \;\big|\; \xi_1\big) = \mathbb{E}\big(F(\xi_1,\xi_3) \;\big|\; \xi_1\big)$. Therefore,

$$V_2 = \operatorname{var}\Big[\mathbb{E}\big(F(\xi_1,\xi_2) \;\big|\; \xi_1\big)\Big] \geq 0\,.$$

Thus, if $\mathbb{E}\big(F(\xi_1,\xi_2) \;\big|\; \xi_1\big)$ is not almost surely constant (or equivalently not almost surely zero), we have $V_2 > 0$. In this case, the U-statistic is called *non-degenerate*.

The dependence structure of the family $\big(F(\xi_i,\xi_j)\big)_{1\leq i<j\leq n}$ fits the graph on the vertex set $\{(i,j)\,;\, 1 \leq i < j \leq n\}$, where vertices (i,j) and (k,l) are adjacent if $\{i,j\} \cap \{k,l\} \neq \emptyset$: taking families $(i_\alpha, j_\alpha)_{\alpha\in A}$ and $(k_\beta, l_\beta)_{\beta\in B}$, such that $\{i_\alpha, j_\alpha\} \cap \{k_\beta, l_\beta\} = \emptyset$ for all α and β, the sets $\{i_\alpha, j_\alpha\,;\, \alpha \in A\}$ and $\{k_\beta, l_\beta\,;\, \beta \in B\}$ must be disjoint. Consequently, we may take $D = 2n - 3$.

Rescaling and applying Theorem 3.3.2, we obtain

$$d_{\mathrm{W}}\left(\mathscr{L}\left(\frac{U_n}{\sigma_n}\right), \mathcal{N}(0,1)\right) \leq \frac{n(n-1)(28n^2-48n+59)}{2\left(\frac{n(n-1)}{2}\,V_1 + n(n-1)(n-2)\,V_2\right)^{3/2}}\,\mathbb{E}\big|F(\xi_1,\xi_2)\big|^3\,.$$

Notice that if the statistic is non-degenerate, then the rate of convergence is again $1/\sqrt{n}$, like for independent random variables (see Example 3.2.5).

If the statistic is degenerate, the distributions may not converge to the standard normal. A typical example is if F is of the form $F(x,y) = G(x)\,G(y)$. In this case, we have

$$U_n = \frac{1}{2}\left(\sum_{i=1}^n G(\xi_i)\right)^2 - \frac{1}{2}\sum_{i=1}^n \big(G(\xi_i)\big)^2\,. \tag{3.3.1}$$

Observe that $\mathbb{E}\big[F(\xi_1,\xi_2)\big] = \big(\mathbb{E}\big[G(\xi_1)\big]\big)^2 = 0$, so that $\mathbb{E}\big[G(\xi_1)\big] = 0$ by our assumption. Letting $\tau_1^2 := \operatorname{var}\big[G(\xi_1)\big]$, we find that the first term in the right hand side of (3.3.1) has expectation $n\tau_1^2$ and variance of order n^2, while the second term has expectation $n\tau_1^2$ and variance of order n. By the central limit theorem, the distribution of the first term divided by n approaches the chi squared distribution with one degree of freedom, scaled by a constant factor. Therefore, the distributions of $U_n/\sqrt{n}$ converge to the centred version of that distribution.

3.3.2 Random permutations

Let $\mathbf{A} = [a(i,j)]_{i,j}$ be a $n \times n$ matrix of real numbers. Consider the statistic

$$W = \sum_{i=1}^n a\big(i, \Pi(i)\big) = \sum_{i=1}^n \sum_{j=1}^n a(i,j)\,\mathbf{1}\big(\Pi(i) = j\big)\,,$$

where Π is a uniformly distributed random permutation of $\{1, 2, \ldots, n\}$. Observe that W changes just for a constant if we add a constant value to all entries of a particular row or column. Subtracting the averages, we can make all rows to have sum zero. Doing the same with the columns, observe that the row sums still remain zero. Therefore, we can assume without for generality that

$$\sum_{j=1}^{n} a(i,j) = 0 \quad \text{for all } i \qquad \text{and} \qquad \sum_{i=1}^{n} a(i,j) = 0 \quad \text{for all } j. \tag{3.3.2}$$

In this case, of course, $\mathbb{E}\, W = 0$. Now compute

$$\operatorname{var}(W) = \mathbb{E}(W^2) = \sum_{1\le i,j\le n} \sum_{1\le i',j'\le n} a(i,j)\, a(k,l)\, \mathbb{P}\big(\Pi(i) = j,\ \Pi(k) = l\big)\,.$$

Noting that

$$\mathbb{P}\big(\Pi(i) = j,\ \Pi(k) = l\big) = \begin{cases} \frac{1}{n} & ;\ i = k, j = l \\ \frac{1}{n(n-1)} & ;\ i \ne k, j \ne l\,, \\ 0 & ;\ \text{otherwise}\,, \end{cases}$$

we further compute

$$\begin{aligned} \operatorname{var}(W) &= \frac{1}{n} \sum_{1\le i,j\le n} a(i,j)^2 + \frac{1}{n(n-1)} \sum_{1\le i,j\le n} \sum_{1\le k,l\le n; k\ne i, l\ne j} a(i,j)\, a(k,l) \\ &= \frac{1}{n} \sum_{1\le i,j\le n} a(i,j)^2 + \frac{1}{n(n-1)} \sum_{1\le i,j\le n} \sum_{1\le k,l\le n} a(i,j)\, a(k,l) \\ &\quad - \frac{1}{n(n-1)} \sum_{1\le i,j\le n} \sum_{1\le l\le n} a(i,j)\, a(i,l) \\ &\quad - \frac{1}{n(n-1)} \sum_{1\le i,j\le n} \sum_{1\le k\le n} a(i,j)\, a(k,j) \\ &\quad + \frac{1}{n(n-1)} \sum_{1\le i,j\le n} a(i,j)^2 \\ &= \frac{1}{n-1} \sum_{1\le i,j\le n} a(i,j)^2\,. \end{aligned}$$

In the sequel, we shall assume (3.3.2) **and** $\operatorname{var}(W) = 1$**.**

To construct decompositions from Proposition 3.1.1, we introduce the concept of simple random relocation.

Definition 3.3.5. Let $A \subseteq M$ be finite sets. A *simple random relocation* of the set A within the set M is a random permutation T_A of the set M, which acts as follows:

- The elements of the set A are mapped to any elements of M uniformly at random.
- Given the latter and denoting by B the image of A under T_A, all elements of $B \setminus A$ are mapped to the elements of $A \setminus B$ uniformly at random.

- The other elements are left unchanged.

Proposition 3.3.6. *Let $A \subseteq M$ be finite sets and let T_A be a simple random relocation of A within M. Take a uniformly distributed random permutation Π of M, independent of T_A. Then $\Pi \circ T_A^{-1}$ is also uniformly distributed and is independent of the restriction of Π to A.*

Proof. Let $A = \{i_1, \dots, i_r\}$. What we need to prove is that given $\Pi(i_1) = j_1, \dots, \Pi(i_r) = j_r$, $\Pi \circ T_A^{-1}$ is uniformly distributed. Suppose that $T_A(i_1) = k_1, \dots, T_A(i_r) = k_r$, and observe that the map $\sigma \mapsto \sigma \circ T_A^{-1}$ is a one-to-one correspondence between the set of permutations σ with $\sigma(i_1) = j_1, \dots, \sigma(i_r) = j_r$ and the set of permutations σ with $\sigma(k_1) = j_1, \dots, \sigma(k_r) = j_r$. Therefore, given $T_A(i_1) = k_1, \dots, T_A(i_r) = k_r, \Pi(i_1) = j_1, \dots, \Pi(i_r) = j_r$, the random permutation $\sigma \circ T_A^{-1}$ is uniformly distributed over all permutations σ with $\sigma(k_1) = j_1, \dots, \sigma(j_r) = j_r$. Noting that the r-tuple $\big(T_A(i_1), \dots, T_A(i_r)\big)$ is uniformly distributed over all possible elements of M^r with distinct elements, the proof is complete. □

Now set

$$\mathscr{I}_i := \{1, 2, \dots, n\}\,, \qquad X_i := a\big(i, \Pi(i)\big)\,,$$

noting that $\mathbb{E}\, X_i = 0$ for all i. Next, take a family of simple random relocations T_A, $A \subseteq \{1, 2, \dots, n\}$ of sets A within $\{1, 2, \dots, n\}$, which are all independent of Π. By Proposition 3.3.6, the random variable

$$W_i := \sum_{j=1}^{N} a\big(j, \Pi(T_{\{i\}}^{-1}(j))\big) = \sum_{j=1}^{N} a\big(T_{\{i\}}(j), \Pi(j)\big)\,, \tag{3.3.3}$$

is independent of X_i for each i, so that we can set

$$\mathscr{I}_i := \{1, 2, \dots, n\}\,, \qquad X_{ij} := a\big(j, \Pi(j)\big) - a\big(T_{\{i\}}(j), \Pi(j)\big)\,, \qquad R_i := \sum_{j \in \mathscr{I}_i} X_{ij}\,,$$

noting that $R_i = W - W_i$.

The pair (X_i, X_{ij}) is uniquely determined by $T_{\{i\}}$ and the restriction of Π to $\{i, j\}$. Thus, take another family of simple random relocations T'_A, $A \subseteq \{1, 2, \dots, n\}$, which is independent of all other random variables. Similarly as before, put

$$W_{ij} := \sum_{k=1}^{N} a\big(k, \Pi(T'^{-1}_{\{i,j\}}(k))\big) = \sum_{j=1}^{N} a\big(T'_{\{i,j\}}(k), \Pi(k)\big) \tag{3.3.4}$$

and by Proposition 3.3.6, W_{ij} is independent of the pair (X_i, X_{ij}).

To apply Corollary 3.2.4, we need to bound the sums

$$\begin{gathered}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\big[|X_i|\,|X_{ij}|\,|R_i + R_{ij}|\big]\,, \qquad \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\big[|X_i|\,|X_{ij}|\,|R_{ij}|\big]\,, \\ \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\big[|X_i|\,|X_{ij}|\big]\,\mathbb{E}\,|R_i + R_{ij}|\,.\end{gathered} \tag{3.3.5}$$

As T_A preserves all points outside $A \cup T_A(A) = A \cup T_A^{-1}(A)$ (and similarly T'_A), we have

$$\begin{aligned}
|X_{ij}| &\le \mathbf{1}\Big[j \in \big\{i, T_{\{i\}}^{-1}(i)\big\}\Big]\Big(\big|a\big(j,\Pi(j)\big)\big| + \big|a\big(T_{\{i\}}(j),\Pi(j)\big)\big|\Big)\,,\\
|R_i + R_{ij}| &\le \sum_{k=1}^{n} \mathbf{1}\Big[k \in \big\{i, j, T'^{-1}_{\{i,j\}}(i), T'^{-1}_{\{i,j\}}(j)\big\}\Big]\Big(\big|a\big(k,\Pi(k)\big)\big| + \big|a\big(T'_{\{i,j\}}(k),\Pi(k)\big)\big|\Big)\,,\\
|R_{ij}| &\le \sum_{k=1}^{n} \mathbf{1}\Big[k \in \big\{i, j, T_{\{i\}}^{-1}(i), T'^{-1}_{\{i,j\}}(i), T'^{-1}_{\{i,j\}}(j)\big\}\Big]\\
&\qquad\qquad \times \Big(\big|a\big(T_{\{i\}}(k),\Pi(k)\big)\big| + \big|a\big(T'_{\{i,j\}}(k),\Pi(k)\big)\big|\Big)\,.
\end{aligned} \tag{3.3.6}$$

The resulting bounds on the sums in (3.3.5) can be expressed in terms of auxiliary random indices and permutations: let I be uniformly distributed over $\{1, 2, \dots, n\}$ and independent of all other random elements. Letting

$$\begin{gathered}
J_1 := I\,, \qquad J_2 := T_{\{I\}}^{-1}(I)\,;\\
J_{r1} := J_r\,, \qquad J_{r2} := T_{\{I\}}(J_r)\,;\\
K_{r1} := I\,, \qquad K_{r2} := J_r\,, \qquad K_{r3} := T'^{-1}_{\{I,J_r\}}(I)\,, \qquad K_{r4} := T'^{-1}_{\{I,J_r\}}(J_r)\,;\\
K_{rs1} := K_{rs}\,, \qquad J_{rs2} := T_{\{I,J_r\}}(K_{rs})\,,
\end{gathered}$$

we find that

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\big[|X_i|\,|X_{ij}|\,|R_i + R_{ij}|\big]\\
&\qquad \le n\,\mathbb{E}\Bigg[\big|a\big(I,\Pi(I)\big)\big| \sum_{r=1}^{2}\sum_{u=1}^{2}\big|a\big(J_{ru},\Pi(J_r)\big)\big| \sum_{s=1}^{4}\sum_{v=1}^{2}\big|a\big(K_{rsv},\Pi(K_{rs})\big)\big|\Bigg]\,.
\end{aligned}$$

Applying the inequality between the arithmetic and the geometric mean, we obtain

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\big[|X_i|\,|X_{ij}|\,|R_i + R_{ij}|\big]\\
&\qquad \le \frac{n}{3}\sum_{r=1}^{2}\sum_{u=1}^{2}\sum_{s=1}^{4}\sum_{v=1}^{2}\Big[\mathbb{E}\big|a\big(I,\Pi(I)\big)\big|^3 + \mathbb{E}\big|a\big(J_{ru},\Pi(J_r)\big)\big|^3 + \mathbb{E}\big|a\big(K_{rsv},\Pi(K_{rs})\big)\big|^3\Big]\,.
\end{aligned}$$

Using independence, we find that each random index L being equal to I, J_{ru} or K_{rsv} is uniformly distributed over $\{1, 2, \dots, n\}$. Moreover, since Π is independent of all these random indices, $\Pi(L)$ is independent of L and also uniformly distributed over $\{1, 2, \dots, n\}$. This leads to the bound

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\big[|X_i|\,|X_{ij}|\,|R_i + R_{ij}|\big] \le \frac{32}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} |a(i,j)|^3\,.$$

Similarly, introducing $K_{r5} := T'^{-1}_{\{I\}}(I)$, we estimate

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\big[|X_i|\,|X_{ij}|\,|R_{ij}|\big] \\
&\qquad\le n\,\mathbb{E}\Big[\big|a\big(I,\Pi(I)\big)\big|\sum_{r=1}^{2}\sum_{u=1}^{2}\big|a\big(J_{ru},\Pi(J_r)\big)\big|\sum_{s=1}^{5}\sum_{v=1}^{2}\big|a\big(K_{rsv},\Pi(K_{rs})\big)\big|\Big] \\
&\qquad\le \frac{40}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|a(i,j)|^3\,.
\end{aligned}$$

Finally, to estimate the last sum in (3.3.5), we introduce another random permutation Π', which is uniformly distributed and independent of all other random variables. Similarly as before, we obtain

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\big[|X_i|\,|X_{ij}|\big]\,\mathbb{E}\,|R_i+R_{ij}| \\
&\qquad\le n\,\mathbb{E}\Big[\big|a\big(I,\Pi(I)\big)\big|\sum_{r=1}^{2}\sum_{u=1}^{2}\big|a\big(J_{ru},\Pi(J_r)\big)\big|\sum_{s=1}^{4}\sum_{v=1}^{2}\big|a\big(K_{rsv},\Pi'(K_{rs})\big)\big|\Big] \\
&\qquad\le \frac{32}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|a(i,j)|^3\,.
\end{aligned}$$

Collecting all together and applying Corollary 3.2.4, we obtain the following result:

Proposition 3.3.7. *Let $a(i,j)$, $1\le i,j\le n$, be real numbers, such that*

$$\sum_{j=1}^{n}a(i,j)=0\quad\text{for all } i\qquad\text{and}\qquad\sum_{i=1}^{n}a(i,j)=0\quad\text{for all } j.$$

Take a uniformly distributed random permutation Π of the set $\{1,2,\dots,n\}$ and let

$$W:=\sum_{i=1}^{n}a\big(i,\Pi(i)\big)\,.$$

If $\operatorname{var}(W)=1$, *then we have*

$$d_{\mathrm{W}}\big(\mathscr{L}(W),\,\mathcal{N}(0,1)\big)\le\frac{136}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|a(i,j)|^3\,.$$

□

Remark 3.3.8. More careful computations along with a more sophisticated version of Corollary 3.2.4 would yield a better constant. Moreover, as mentioned in Example 2.4.4, it is plausible that a sum of m independent random variables can be obtained as a limit of statistics W defined as above, where $a(i,j)=0$ for $i>m$, m is fixed and n tends to infinity. It is possible to derive a bound in the normal approximation which approaches the bound for sums of independent random variables as stated in Example 3.3.3.

3.4 The Berry–Esseen theorem

The celebrated Berry–Esseen theorem bounds the error in the central limit theorem for independent and identically distributed random variables in terms of the *Kolmogorov distance*:

$$d_{\mathrm{K}}(\mu,\nu) := \sum_{a\in\mathbb{R}} \Big|\mu\big((-\infty,a]\big) - \nu\big((-\infty,a]\big)\Big| .$$

Theorem 3.4.1 (Berry [7], Esseen [16]). *Let $\xi_1, \xi_2, \dots$ be independent and identically distributed random variables with $\mathbb{E}\,\xi_1 = 0$, $\operatorname{var}(\xi_1) = 1$ and $\mathbb{E}\,|\xi_1|^3 < \infty$. Letting*

$$W^{(n)} := \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{\sqrt{n}} ,$$

we have

$$d_{\mathrm{K}}\big(\mathscr{L}\big(W^{(n)}\big),\, \mathcal{N}(0,1)\big) \le \frac{C\,\mathbb{E}\,|\xi_1|^3}{\sqrt{n}} ,$$

where C is a universal constant.

Remark 3.4.2. The calculation of the constant C has a long history. In 1941, Berry [7] claimed that the result holds with $C = 1.88$, but his calculations turned out not to be entirely correct (see Hsu [21]). In 1945, Esseen [16] proved the result with $C = 7.59$. Over the next decades, the constant was significantly improved. The best value obtained so far seem to be $C = 0.4748$, obtained in 2011 by Shevtsova [28] (the same author even claims $C = 0.469$ in her paper [29], but provides no proof). Shevtsova's bound is not far from optimal: a lower bound $\frac{\sqrt{10}+3}{6\sqrt{2\pi}} > 0.4097$ on the constant C was derived in 1956 by Esseen [17]. For more details on the history of calculation of C, see Korolev and Shevtsova [22].

Remark 3.4.3. Example 3.3.3 provides a bound in the Wasserstein distance:

$$d_{\mathrm{W}}\big(\mathscr{L}\big(W^{(n)}\big),\, \mathcal{N}(0,1)\big) \le \frac{3\,\mathbb{E}\,|\xi_1|^3}{\sqrt{n}} ,$$

which, combined with (A.2.7), gives

$$d_{\mathrm{W}}\big(\mathscr{L}\big(W^{(n)}\big),\, \mathcal{N}(0,1)\big) \le \frac{\sqrt{6}}{\sqrt[4]{2\pi}} \frac{\sqrt{\mathbb{E}\,|\xi_1|^3}}{\sqrt[4]{n}} .$$

However, the latter bound does not preserve the rate of convergence.

Here, we prove Theorem 3.4.1 by Stein's method. Unfortunately, the constant will be far from optimal. However, the main advantage of Stein's method is that it can be readily applied to sums of dependent random variables, where most other methods do not seem work. In particular, most of the improvements of the constant in the Berry–Esseen theorem (including the improvement by Shevtsova [28]) are proved by the method of characteristic functions. For sums of independent random variables, characteristic functions satisfy the multiplication formula, which has no straightforward extension to, let's say, local dependence.

Thus, the proof of the Berry–Esseen theorem can serve as a guideline how to extend the result to sums of dependent random variables. Here, we shall not do the latter. In fact, bounds of correct order in terms of the Kolmogorov metric are much harder to obtain than for the Wasserstein metric. In particular, it seems to be very difficult to derive a result being as general as Corollary 3.2.4. Instead, more special results have been derived. Bolthausen [8] proves a Berry–Esseen type result for random permutations. Chen and Shao [12] prove a result for local dependence, but that result does not yield the counterpart of Theorem 3.3.2 for local dependence. One can do more under the assumption of higher moments: see Chen, Goldstein and Röllin [10] and references therein. The assumption of boundedness admits even more general and cleaner results: see Dembo and Rinott [15], Goldstein [19] and Raič [26].

In order to derive the Berry–Esseen theorem by Stein's method, it is beneficial to consider functions with bounded total variation in view of Section A.3.

Theorem 3.4.4. *Consider a function* $h\colon \mathbb{R} \to \mathbb{R}$.

(1) *If* h *has bounded variation, then the function* f *defined by* (3.2.4) *is an almost-everywhere solution to the Stein equation* (3.2.1)*, which satisfies* $V(f') \le 2\,V(h)$.

(2) *If* h *is absolutely continuous and* h' *has bounded variation, the function* f *defined by* (3.2.4) *is a classical solution to the Stein equation* (3.2.1)*. Moreover,* f' *is absolutely continuous and* $V(f'') \le 2\,V(h')$.

Proof.

Part (1). First recall that h is bounded by Remark B.2.2 and measurable by Corollary B.2.5. Therefore, $\mathcal{N}|h| < \infty$. By Proposition 3.2.1, there exists an almost-everywhere solution f to the Stein equation, which means that f is absolutely continuous. Recalling (3.2.6), there is an almost-everywhere derivative of f given by

$$f'(w) = h(w) - \psi'(-w) \int_{-\infty}^{w} h(x)\,\phi(x)\,\mathrm{d}x - \psi'(w) \int_{w}^{\infty} h(x)\,\phi(x)\,\mathrm{d}x\,. \tag{3.4.1}$$

We shall derive a formula similar to (3.2.7), which was derived from (3.2.6) by integration by parts, integrating ϕ and differentiating h. This can be done of h is absolutely continuous, which is not an assumption in our case. However, we can use the (improper) *Riemann–Stieltjes integral*. By assumption, h has bounded variation. Letting $\Phi^-(y) := \Phi(-y)$, Φ and Φ^- are absolutely continuous. By Proposition B.3.8, we can then rewrite (3.4.1) as

$$f'(w) = h(w) - \psi'(-w) \int_{-\infty}^{w} h(x)\,\mathrm{d}\Phi(x) + \psi'(w) \int_{w}^{\infty} h(x)\,\mathrm{d}\Phi^-(x)\,.$$

By the integration by parts formula (Proposition B.3.6) and noting that

$\lim_{w\to-\infty} h(w)\,\Phi(w) = \lim_{w\to\infty} h(w)\,\Phi(-w) = 0$, we have

$$\begin{aligned} f'(w) &= h(w) - \psi'(-w)\,h(w)\,\Phi(w) + \psi'(-w)\int_{-\infty}^{w} \Phi(y)\,\mathrm{d}h(y) \\ &\quad - \psi'(w)\,h(w)\,\Phi(-w) - \psi'(w)\int_{w}^{\infty} \Phi(-y)\,\mathrm{d}h(y) \\ &= \psi'(-w)\int_{-\infty}^{w} \Phi(y)\,\mathrm{d}h(y) - \psi'(w)\int_{w}^{\infty} \Phi(-y)\,\mathrm{d}h(y)\,, \end{aligned}$$

where the last equality is due to Proposition C.2.3 for $r = 1$.

Let Λ_h be the signed measure associated to h in view of Definition B.4.9. By part (2) of Proposition B.4.11, we have $\int_a^w \Phi(y)\,\mathrm{d}h(y) = \int_{(a,w]} \Phi\,\mathrm{d}\Lambda_h$ and $\int_w^b \Phi(-y)\,\mathrm{d}h(y) = \int_{(w,b]} \Phi^-\,\mathrm{d}\Lambda_h$ for all $a \le w \le b$. Since Φ is bounded on $(-\infty, w]$ and Φ^- is bounded on (w,∞), we may take the limit in a and b, leading to $\int_{-\infty}^w \Phi(y)\,\mathrm{d}h(y) = \int_{(-\infty,w]} \Phi\,\mathrm{d}\Lambda_h$ and $\int_w^b \Phi(-y)\,\mathrm{d}h(y) = \int_{(w,\infty)} \Phi^-\,\mathrm{d}\Lambda_h$. We may rewrite this as

$$f'(w) = \int_{-\infty}^{\infty} F(w,y)\,\Lambda_h(\mathrm{d}y)\,,$$

where

$$F(w,y) = \begin{cases} -\psi'(w)\,\Phi(-y) & ;\ w < y \\ \psi'(-w)\,\Phi(y) & ;\ w \ge y\,. \end{cases}$$

Letting $F_y(w) := F(w,y)$, it is left to the reader as an exercise to show that $\sum_{k=1}^n \big|F_y(w_k) - F_y(w_{k-1})\big| \le 2\big(\psi'(y)\,\Phi(-y) + \psi'(-y)\,\Phi(y)\big)$ for every sequence $w_0 \le w_1 \le \cdots \le w_n$ (use Proposition C.1.5). Therefore, $V(F_y) \le 2\big(\psi'(y)\,\Phi(-y) + \psi'(-y)\,\Phi(y)\big)$. Moreover, by Proposition C.1.6, we have $\lim_{t\to\infty}\big[\big|F_y(-t) - F_y\big(w - \frac{1}{t}\big)\big| + \big|F_y\big(w - \frac{1}{t}\big)\big| - \big|F_t(y)\big| + \big|F_y(y) - F_y(t)\big|\big] = 2\big(\psi'(y)\,\Phi(-y) + \psi'(-y)\,\Phi(y)\big)$. By Proposition C.2.3, we have $\psi'(y)\,\Phi(-y) + \psi'(-y)\,\Phi(y) = 1$, so that $V(F_y) = 2$ for all y. Finally, applying part (2) of Proposition B.2.7 along with part (1) of Proposition B.4.11, we estimate $V(f') \le \int_{-\infty}^{\infty} V(F_y)\,|\Lambda_h|(\mathrm{d}y) = 2\|\Lambda_h\| = 2\,V(h)$, proving part (1).

Part (2). We proceed similarly as in part (1). First recall that h' is bounded by Remark B.2.2. Therefore, h is of linear growth and $\mathcal{N}|h| < \infty$. By Proposition 3.2.1, there exists a classical solution f to the Stein equation, such that f' is absolutely continuous and, recalling (3.2.8),

$$f''(w) = h'(w) - \psi''(-w)\int_{-\infty}^{w} h'(y)\,\Phi(y)\,\mathrm{d}y - \psi''(w)\int_{w}^{\infty} h'(y)\,\Phi(-y)\,\mathrm{d}y\,. \tag{3.4.2}$$

Again, we rewrite this formula in terms of the Riemann–Stieltjes integral. By assumtion, h' has finite total variation. The functions $\Phi_2(y) := \int_{-\infty}^y \Phi(t)\,\mathrm{d}t$ $\Phi_2^-(y) := \Phi_2(-y)$ are absolutely continuous. By Proposition B.3.8, we have

$$f''(w) = h'(w) - \psi''(-w)\int_{-\infty}^{w} h'(y)\,\mathrm{d}\Phi_2(y) + \psi''(w)\int_{w}^{\infty} h'(y)\,\mathrm{d}\Phi_2^-(y)\,.$$

By the integration by parts formula (Proposition B.3.6) and noting that $\lim_{w\to-\infty} h'(w)\,\Phi_2(w) = \lim_{w\to\infty} h'(w)\,\Phi_2(-w) = 0$ by Corollary C.2.2, we have

$$\begin{aligned} f''(w) &= h'(w) - \psi''(-w)\,h'(w)\,\Phi_2(w) + \psi''(-w)\int_{-\infty}^{w}\Phi_2(y)\,\mathrm{d}h'(y) \\ &\quad - \psi''(w)\,h'(w)\,\Phi_2(-w) - \psi''(w)\int_{w}^{\infty}\Phi_2(-y)\,\mathrm{d}h'(y) \\ &= \psi''(-w)\int_{-\infty}^{w}\Phi_2(y)\,\mathrm{d}h'(y) - \psi''(w)\int_{w}^{\infty}\Phi_2(-y)\,\mathrm{d}h'(y)\,, \end{aligned}$$

where the last equality is due to Proposition C.2.3 for $r = 2$. Similarly as in the proof of part (1), we can rewrite this as

$$f''(w) = \int_{-\infty}^{\infty} G(w,y)\,\Lambda_{h'}(\mathrm{d}y)\,,$$

where

$$G(w,y) = \begin{cases} -\psi''(w)\,\Phi_2(-y) & ;\ w < y \\ \psi''(-w)\,\Phi_2(y) & ;\ w \geq y\,. \end{cases}$$

Letting $G_y(w) := G(w,y)$, we again find that $V(G_y) = 2\big(\psi''(y)\,\Phi_2(-y) + \psi''(-y)\,\Phi_2(y)\big) = 2$, with the last equality due to Proposition C.2.3 for $r = 2$. Finally, applying part (2) of Proposition B.2.7 along with part (1) of Proposition B.4.11, we estimate $V(f'') \leq \int_{-\infty}^{\infty} V(F_y)\,|\Lambda_{h'}|(\mathrm{d}y) = 2\|\Lambda_{h'}\| = 2\,V(h')$, completing the proof. □

Proof of Theorem 3.4.1. Sums of independent random variables are clearly a very special case of decompositions of Barbour, Karoński and Ruciński introduced in Section 3.1. Letting

$$\mathscr{I} := \{1,2,\ldots,n\}\,, \qquad \mathscr{J}_i := \{0\}\,,$$
$$X_i^{(n)} := R_i^{(n)} := X_{i0}^{(n)} := \frac{\xi_i}{\sqrt{n}}\,, \qquad W_i^{(0)} := W_{i0}^{(n)} := W^{(n)} - X_i^{(n)}\,, \qquad R_{i0}^{(n)} := 0\,,$$

all conditions specified in Section 3.1 are fulfilled and (3.1.5) reduces to

$$\begin{aligned} \mathbb{E}\big[f'(W) - f(W)W\big] = \sum_{i=1}^{n}\mathbb{E}\Big[&f''\big(W_i^{(n)} + \theta_2 X_i^{(n)}\big)X_i^{(n)}\,\mathbb{E}\big(\big(X_i^{(n)}\big)^2\big) \\ &- \theta_1\,f''\big(W_i^{(n)} + \theta_1\theta_2 X_i^{(n)}\big)\big(X_i^{(n)}\big)^3\Big]\,, \end{aligned} \tag{3.4.3}$$

where θ_1 and θ_2 are uniformly distributed over $[0,1]$ and independent of each other as well as of $X_1^{(n)},\ldots,X_n^{(n)}$. This is true for all functions f with $M_2(f) < \infty$ (see Proposition 3.1.1).

Denoting the indicators of half-lines by

$$h_a(w) := \begin{cases} 1 & ;\ w \leq a \\ 0 & ;\ x > a\,, \end{cases}$$

the error in the normal approximation in the Kolmogorov metric is expressed as

$$d_{\mathrm{K}}\Big(\mathscr{L}\big(W^{(n)}\big),\, \mathcal{N}(0,1)\Big) = \sup_{a\in\mathbb{R}}\Big|\mathbb{E}\big[h_a\big(W^{(n)}\big)\big] - \mathcal{N}h_a\Big|\,.$$

Unfortunately, the derivative of the solution of the Stein equation (3.2.1) with h_a in place of h is not absolutely continuous (exercise). Therefore, similarly as in the proof of Propositions A.2.9 and A.4.11, we approximate the *step* functions h_a by the *slope* functions

$$h_{a,b}(x) := \begin{cases} 1 & ;\, x \le a \\ \dfrac{b-x}{b-a} & ;\, a \le x \le b \\ 0 & ;\, x \ge b\,, \end{cases}$$

defined for $a < b$. Observe that

$$\begin{aligned} \mathcal{N}h_{a,b} - \mathcal{N}h_a &= \int_a^b \frac{b-x}{b-a}\,\phi(x)\,\mathrm{d}x \le \frac{1}{\sqrt{2\pi}}\int_a^b \frac{b-x}{b-a}\,\mathrm{d}x = \frac{b-a}{2\sqrt{2\pi}}\,, \\ \mathcal{N}h_b - \mathcal{N}h_{a,b} &= \int_a^b \frac{x-a}{b-a}\,\phi(x)\,\mathrm{d}x \le \frac{1}{\sqrt{2\pi}}\int_a^b \frac{x-a}{b-a}\,\mathrm{d}x = \frac{b-a}{2\sqrt{2\pi}}\,. \end{aligned}$$

Therefore, for any $a \in \mathbb{R}$ and $\varepsilon > 0$,

$$\begin{aligned} \mathbb{E}\big[h_a\big(W^{(n)}\big)\big] - \mathcal{N}h_a &\le \mathbb{E}\big[h_{a,a+\varepsilon}\big(W^{(n)}\big)\big] - \mathcal{N}h_{a,a+\varepsilon} + \frac{\varepsilon}{2\sqrt{2\pi}}\,, \\ \mathcal{N}h_a - \mathbb{E}\big[h_a\big(W^{(n)}\big)\big] &\le \mathcal{N}h_{a-\varepsilon,a} - \mathbb{E}\big[h_{a-\varepsilon,a}\big(W^{(n)}\big)\big] + \frac{\varepsilon}{2\sqrt{2\pi}} \end{aligned}$$

and consequently

$$d_{\mathrm{K}}\Big(\mathscr{L}\big(W^{(n)}\big),\, \mathcal{N}(0,1)\Big) \le \sup_{a\in\mathbb{R}}\Big|\mathbb{E}\big[h_{a,a+\varepsilon}\big(W^{(n)}\big)\big] - \mathcal{N}h_{a,a+\varepsilon}\Big| + \frac{\varepsilon}{2\sqrt{2\pi}}\,. \tag{3.4.4}$$

Let $f_{a,b}$ be the solution to the Stein equation (3.2.1) defined by (3.2.4) with $f_{a,b}$ in lace of f and $h_{a,b}$ in place of h. Since $M_1(h_{a,b}) < \infty$, we have $M_1(f_{a,b}) < \infty$ by Theorem 3.2.3, so that (3.4.3) applies with $f_{a,b}$ in place of f. Therefore,

$$\begin{aligned} \mathbb{E}\big[h_{a,a+\varepsilon}\big(W^{(n)}\big)\big] - \mathcal{N}h_{a,a+\varepsilon} = \sum_{i=1}^n \mathbb{E}\Big[& f''_{a,a+\varepsilon}\big(W_i^{(n)} + \theta_2 X_i^{(n)}\big) X_i^{(n)}\, \mathbb{E}\big(\big(X_i^{(n)}\big)^2\big) \\ & - \theta_1\, f''_{a,a+\varepsilon}\big(W_i^{(n)} + \theta_1\theta_2 X_i^{(n)}\big)\big(X_i^{(n)}\big)^3\Big]\,. \end{aligned} \tag{3.4.5}$$

Now fix $x \in \mathbb{R}$ and consider the expectation

$$\mathbb{E}\Big[f''_{a,a+\varepsilon}\big(W_i^{(n)} + x\big)\Big] = \mathbb{E}\left[f''_{a,a+\varepsilon}\left(\sqrt{\frac{n-1}{n}}\,W^{(n-1)} + x\right)\right]. \tag{3.4.6}$$

Denoting by $\mathcal{N}(\mu,\sigma^2)$ the normal distribution with mean μ and variance σ^2, write

$$\left\langle f''_{a,a+\varepsilon}\,,\, \mathcal{N}\left(x, \frac{n-1}{n}\right)\right\rangle = \sqrt{\frac{n}{n-1}}\int_{-\infty}^{\infty} f''_{a,a+\varepsilon}(w)\,\phi\left((w-x)\sqrt{\frac{n}{n-1}}\right)\mathrm{d}w\,.$$

However, since $\int_{-\infty}^{\infty} f''_{a,a+\varepsilon}(w)\,\mathrm{d}w = \lim_{w\to\infty} f'_{a,a+\varepsilon}(w) - \lim_{w\to\infty} f'_{a,a+\varepsilon}(w) = 0$, we also have

$$\left\langle f''_{a,a+\varepsilon}\,,\,\mathcal{N}\!\left(x,\frac{n-1}{n}\right)\right\rangle = \sqrt{\frac{n}{n-1}}\int_{-\infty}^{\infty} f''_{a,a+\varepsilon}(w)\,\left[\phi\left((w-x)\sqrt{\frac{n}{n-1}}\right) - \frac{1}{2\sqrt{2\pi}}\right]\mathrm{d}w\,.$$

Applying by Proposition B.2.6 and part (1) of Theorem 3.4.4, we estimate

$$\begin{aligned}\left|\left\langle f''_{a,a+\varepsilon}\,,\,\mathcal{N}\!\left(x,\frac{n-1}{n}\right)\right\rangle\right| &\le \frac{1}{2\sqrt{2\pi}}\sqrt{\frac{n}{n-1}}\int_{-\infty}^{\infty}|f''_{a,a+\varepsilon}(w)|\,\mathrm{d}w\\ &= \frac{1}{2\sqrt{2\pi}}\sqrt{\frac{n}{n-1}}\,V(f'_{a,a+\varepsilon})\\ &\le \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{n-1}}\,V(h_{a,a+\varepsilon})\\ &= \frac{\sqrt{n}}{\sqrt{2\pi(n-1)}}\,.\end{aligned} \tag{3.4.7}$$

Next, observe that, by Corollary A.3.3 and part (2) of Theorem 3.4.4,

$$\begin{aligned}&\left|\mathbb{E}\left[f''_{a,a+\varepsilon}\!\left(\sqrt{\frac{n-1}{n}}\,W^{(n-1)}+x\right)\right] - \left\langle f''_{a,a+\varepsilon}\,,\,\mathcal{N}\!\left(x,\frac{n-1}{n}\right)\right\rangle\right|\\ &\quad\le V(f''_{a,a+\varepsilon})\;d_{\mathrm{K}}\left(\mathscr{L}\left(\sqrt{\frac{n-1}{n}}\,W^{(n-1)}+x\right),\,\mathcal{N}\!\left(x,\frac{n-1}{n}\right)\right)\\ &\quad= V(f''_{a,a+\varepsilon})\;d_{\mathrm{K}}\Big(\mathscr{L}\big(W^{(n-1)}\big)\,,\,\mathcal{N}(0,1)\Big)\\ &\quad\le 2\,V(h'_{a,a+\varepsilon})\;d_{\mathrm{K}}\Big(\mathscr{L}\big(W^{(n-1)}\big)\,,\,\mathcal{N}(0,1)\Big)\\ &\quad= \frac{4}{\varepsilon}\,d_{\mathrm{K}}\Big(\mathscr{L}\big(W^{(n-1)}\big)\,,\,\mathcal{N}(0,1)\Big)\,.\end{aligned} \tag{3.4.8}$$

Combining (3.4.6), (3.4.7) and (3.4.8), we obtain

$$\left|\mathbb{E}\Big[f''_{a,a+\varepsilon}\big(W_i^{(n)}+x\big)\Big]\right| \le \frac{\sqrt{n}}{\sqrt{2\pi(n-1)}} + \frac{4\delta_{n-1}}{\varepsilon}\,,$$

where

$$\delta_n := d_{\mathrm{K}}\Big(\mathscr{L}\big(W^{(n)}\big)\,,\,\mathcal{N}(0,1)\Big)\,.$$

We also have

$$\left|\mathbb{E}\Big[f''_{a,a+\varepsilon}\big(W_i^{(n)}+\theta_2 X_i^{(n)}\big)\;\Big|\;X_i^{(n)},\theta_1,\theta_2\Big]\right| \le 2\sqrt{\frac{n}{n-1}} + \frac{4\delta_{n-1}}{\varepsilon}\,,$$

$$\left|\mathbb{E}\Big[f''_{a,a+\varepsilon}\big(W_i^{(n)}+\theta_1\theta_2 X_i^{(n)}\big)\;\Big|\;X_i^{(n)},\theta_1,\theta_2\Big]\right| \le 2\sqrt{\frac{n}{n-1}} + \frac{4\delta_{n-1}}{\varepsilon}\,.$$

Plugging into (3.4.5), we obtain

$$\begin{aligned}&\left|\mathbb{E}\big[h_{a,a+\varepsilon}\big(W^{(n)}\big)\big]-\mathcal{N}h_{a,a+\varepsilon}\right|\\&\qquad\le\left(\frac{\sqrt{n}}{\sqrt{2\pi(n-1)}}+\frac{4\delta_{n-1}}{\varepsilon}\right)\sum_{i=1}^{n}\Big(\mathbb{E}\big|X_i^{(n)}\big|\,\mathbb{E}\big(X_i^{(n)}\big)^2+\mathbb{E}\Big(\theta_1\big|X_i^{(n)}\big|^3\Big)\Big]\,.\end{aligned}$$

By Jensen's inequality, we have $\mathbb{E}\big|X_i^{(n)}\big|\le\Big(\mathbb{E}\big|X_i^{(n)}\big|^3\Big)^{1/3}$ and $\mathbb{E}\big(X_i^{(n)}\big)^2\le\Big(\mathbb{E}\big|X_i^{(n)}\big|^3\Big)^{2/3}$, leading to

$$\begin{aligned}\left|\mathbb{E}\big[h_{a,a+\varepsilon}\big(W^{(n)}\big)\big]-\mathcal{N}h_{a,a+\varepsilon}\right|&\le\left(\frac{3}{2}\,\frac{\sqrt{n}}{\sqrt{2\pi(n-1)}}+\frac{6\delta_{n-1}}{\varepsilon}\right)n\,\mathbb{E}\big|X_1^{(n)}\big|^3\\&=\left(\frac{3\sqrt{n}}{\sqrt{8\pi(n-1)}}+\frac{6\delta_{n-1}}{\varepsilon}\right)\frac{\mathbb{E}\,|\xi_1|^3}{\sqrt{n}}\,.\end{aligned}$$

Combining with (3.4.4), we obtain

$$\delta_n\le\left(\frac{3\sqrt{n}}{\sqrt{8\pi(n-1)}}+\frac{6\delta_{n-1}}{\varepsilon}\right)\frac{\mathbb{E}\,|\xi_1|^3}{\sqrt{n}}+\frac{\varepsilon}{2\sqrt{2\pi}}\,.$$

It is easy to check that $\inf_{\varepsilon>0}\big(\frac{a}{\varepsilon}+b\varepsilon\big)=2\sqrt{ab}$ for all $a,b\ge0$. Therefore,

$$\delta_n\le\frac{3\,\mathbb{E}\,|\xi_1|^3}{\sqrt{8\pi(n-1)}}+\sqrt{\frac{12\,\mathbb{E}\,|\xi_1|^3}{\sqrt{2\pi}}\,\frac{\delta_{n-1}}{\sqrt{n}}}\,.$$

Letting $C_n:=\delta_n\sqrt{n}/\,\mathbb{E}\,|\xi_1|^3$, we rewrite this as

$$C_n\le\frac{3\sqrt{n}}{\sqrt{8\pi(n-1)}}+\sqrt{\frac{12}{\sqrt{2\pi}}\sqrt{\frac{n}{n-1}}\,C_{n-1}}\,.$$

Trivially, $\delta_n\le1$ for all n. Next, by Jensen's inequality, we have $\mathbb{E}\,|\xi_1|^3\ge\big(\mathbb{E}\,\xi_1^2\big)^{3/2}=1$. Therefore, $C_n\le\sqrt{n}$ for all n. However, our goal is to bound C_n uniformly in n. Numerical calculations show that

$$\sqrt{n}<\frac{3\sqrt{n}}{\sqrt{8\pi(n-1)}}+\sqrt{\frac{12}{\sqrt{2\pi}}\sqrt{\frac{n}{n-1}}\,\sqrt{n-1}}$$

for all $n\le35$. Therefore, in this case, we can merely say that $C_n\le\sqrt{n}$. Letting

$$C_{35}^*:=\sqrt{35}\,,\qquad C_n^*:=\frac{3\sqrt{n}}{\sqrt{8\pi(n-1)}}+\sqrt{\frac{12}{\sqrt{2\pi}}\sqrt{\frac{n}{n-1}}\,C_{n-1}^*}\,,$$

numerical calculations show

n	C_n^*
35	5.916080
36	5.966365
37	5.987821
38	5.996282
39	5.998944
40	5.999055
41	5.998073

Clearly, $C_n \le C_{35}^*$ for all $n \le 35$ and $C_n \le C_n^*$ for all $n \ge 35$. We have $C_{35}^* < C_{36}^* < C_{37}^* < C_{38}^* < C_{39}^* < C_{40}^*$, so that $C_n \le C_{40}^*$ for all $n \le 40$. Now prove by induction that this remains true for $n \ge 40$. Indeed, for $n \ge 41$, observe that

$$\begin{aligned} C_n &\le \frac{3\sqrt{n}}{\sqrt{8\pi(n-1)}} + \sqrt{\frac{12}{\sqrt{2\pi}}\sqrt{\frac{n}{n-1}}\,C_{n-1}} \\ &\le \frac{3\sqrt{41}}{\sqrt{320\,\pi}} + \sqrt{\frac{12}{\sqrt{2\pi}}\sqrt{\frac{41}{40}}\,C_{40}^*} \\ &< 5.998074 < C_{40}^*\,. \end{aligned}$$

Thus, we have proved the result for $C = C_{40}^* < 6$. □

Remark 3.4.5. For large n, we can do slightly better: one can show that $\limsup_{n\to\infty} C_n \le C_*$, where C^* is the unique solution to the equation

$$C^* = \frac{3}{\sqrt{8\pi}} + \sqrt{\frac{12}{\sqrt{2\pi}}\,C^*}\,.$$

The solution can be computed explicitly as $C^* = \frac{15+\sqrt{216}}{\sqrt{2\pi}} < 5.923683$. Thus, the improvement is still very small comparable to the result of Shevtsova [28], which gives $C = 0.4748$.

As regards Stein's method, Chen and Shao [11] succeed to derive the Berry–Esseen inequality with $C = 4.1$ (however, their constant also applies for sums of non-identically distributed random variables). Even this constant is far away from 0.4748. As already mentioned, the advantage of Stein's method is possibility of extension to dependent summands, but we shall not tackle this issue here.

Appendix A

Convergence of probability measures

Note. This appendix requires a basic knowledge of the theory of metric spaces and topology. For basic definitions related to the latter as well for a slightly deeper insight, the reader is referred to [24] or [30].

We are often interested whether a sequence of probability distributions $(\mu_n)_{n\in\mathbb{N}}$ on a measurable space $(S, \mathscr{S})$ converges to a given probability measure μ. In order to make this precise, we have to endow the space $\mathrm{Pr}(S, \mathscr{S})$, the set of all probability measures on $(S, \mathscr{S})$, with a topology. The latter will be based on *test functions*: the probability measures μ and ν are "close" if the integrals $\langle f\,,\,\mu\rangle$ and $\langle f\,,\,\nu\rangle$ are close for a suitable class of measurable functions f, where we recall from (1.1.1):

$$\langle f\,,\,\mu\rangle := \int f\,\mathrm{d}\mu\,. \tag{A.0.1}$$

Remark A.0.6. If the test functions are unbounded, they cannot test all probability probability measures. In this case, $\mathrm{Pr}(S, \mathscr{S})$ should be replaced by a suitable subspace.

Throughout this appendix, $(S, \mathscr{S})$ will denote a measurable space, while $\mathscr{M}$ will denote a subspace of $\mathrm{Pr}(S, \mathscr{S})$.

A.1 Metrics based on test functions

One of the ways to construct a topology on $\mathscr{M}$ from a class of test functions $\mathscr{F}$ is to define the following metric:

$$d_{\mathscr{F}}(\mu,\nu) = \sup_{f\in\mathscr{F}}\big|\langle f\,,\,\nu\rangle - \langle f\,,\,\mu\rangle\big|\,. \tag{A.1.1}$$

Of course, we must check whether the right hand side is well defined and whether it represents a metric. Firstly, $\langle |f|\,,\,\mu\rangle$ must be finite for all $\mu\in\mathscr{M}$ and $f\in\mathscr{F}$. However, this does not imply that $d(\mu,\nu) < \infty$. Suppose that the latter is true. In this case, the symmetry is obvious and the triangle inequality is easy to check, too. It remains to check

that $d(\mu,\nu)=0$ implies $\mu=\nu$. This is obviously true if the class $\mathscr{F}$ is rich enough. The proof of the following assertion is easy and is therefore left to the reader:

Proposition A.1.1. *Let $d_{\mathscr{F}}$ be as in (A.1.1) and let $d(\mu,\nu)<\infty$ for any μ and ν. Then $d_{\mathscr{F}}$ is a metric if and only if the test functions from $\mathscr{F}$ separate the probability measures from $\mathscr{M}$, that is, when for any two probability measures $\mu\neq\nu\in\mathscr{M}$, there exists a function $f\in\mathscr{F}$, such that $\langle f\,,\,\mu\rangle\neq\langle f\,,\,\nu\rangle$.* □

Example A.1.2. Letting $\mathscr{F}$ be the class of indicators of all measurable sets, we obtain the *total variation metric*, which will be denoted by d_{TV}. Thus, for example, Proposition 2.2.2 can be rewritten as

$$d_{\mathrm{TV}}\big(\mathscr{L}(W),\,\mathrm{Po}(\lambda)\big)\leq\frac{1-e^{-\lambda}}{\lambda}\sum_{i\in\mathscr{I}}p_i^2\,.$$

where $\mathscr{L}(W)$ denotes the distribution (law) of W, that is, $\mathscr{L}(W)(A)=\mathbb{P}(W\in A)$. Clearly, we can take $\mathscr{M}=\Pr(S,\mathscr{S})$ and the class of indicators of all measurable sets separates all probability measures *by definition*, so that d_{TV} is a metric.

Remark A.1.3. The total variation metric typically works well in discrete spaces. Otherwise, it is usually too strong. As an example, consider a sequence $x_1,x_2,\ldots$ of points in a metric space, which converges to x. If all the points x_n are different from x, then $d_{\mathrm{TV}}(\delta_{x_n},\delta_x)=1$ for all n, so that the sequence of Dirac measures $\delta_{x_1},\delta_{x_2},\ldots$ does not converge to δ_x.

Example A.1.4. Letting $\mathscr{F}$ be the class of indicators of all half-lines $(-\infty,a]$ on the real line endowed with the Borel σ-algebra $\mathscr{B}(\mathbb{R})$, we obtain the *Kolmogorov metric*, which will be denoted by d_{K}. Again, we can take $\mathscr{M}=\Pr\big(\mathbb{R},\mathscr{B}(\mathbb{R})\big)$. We claim that $\mathscr{F}$ separates all probability measures. To show this, take Borel probability measures μ and ν which agree on $\mathscr{F}$, that is, on all all half-lines $(-\infty,a]$. The family of all half-lines is closed under intersections and generates the Borel σ-algebra. By Theorem A.1.6 below, μ and ν agree on all Borel sets, so that $\mu=\nu$. Therefore, $\mathscr{F}$ separates all probability measures, so that d_{K} is also a metric. Clearly, $d_{\mathrm{K}}\leq d_{\mathrm{TV}}$, so that d_{K} is weaker than d_{TV}.

Definition A.1.5. A *π-system* is a collection of sets which is closed under finite intersections.

The following result is a consequence of Dynkin's π–λ theorem (see, for example, Corollary 1.6.3 of Cohn [13]).

Theorem A.1.6. *If two probability measures agree on a π-system which generates a σ-algebra $\mathscr{S}$, then they agree on $\mathscr{S}$.* □

The fact that the set of all half-lines $(-\infty,a]$ separates probability measures can be generalized in the following way:

Proposition A.1.7. *Let $\mathscr{F}$ be a class of measurable functions on $(S,\mathscr{S})$ and let $\mathscr{P}$ be a π-system which generates $\mathscr{S}$. For each $A\in\mathscr{P}$, suppose that there exists a uniformly bounded sequence of functions $f_n\in\operatorname{span}\mathscr{F}$ which converges pointwise to the indicator $\mathbf{1}_A$ of a set A. Then $\mathscr{F}$ separates all probability measures from $\Pr(S,\mathscr{S})$.*

PROOF. Let $\mu, \nu \in \Pr(S, \mathscr{S})$ be such that $\langle f\,,\,\mu\rangle = \langle f\,,\,\nu\rangle$ for all $f \in \mathscr{F}$. Thanks to linearity, this is also true for all $f \in \operatorname{span}\mathscr{F}$. From the dominated convergence theorem, it immediately follows that $\mu(A) = \nu(A)$ for all $A \in \mathscr{P}$. By Theorem A.1.6, $\mu(A) = \nu(A)$ for all $A \in \mathscr{S}$, so that $\mu = \nu$. □

Remark A.1.8. Again, if $x_1, x_2, \ldots$ is a sequence of real numbers converging to x and all numbers x_n are different from x, we still have $d_{\mathrm{K}}(\delta_{x_n}, \delta_x) = 1$ for all n, so that the sequence of Dirac measures $\delta_{x_1}, \delta_{x_2}, \ldots$ does not converge to δ_x in the Kolmogorov metric either.

Now consider another example: take $0 < p < 1$ and let $X_n \sim \mathrm{Bin}(n, p)$, $n \in \mathbb{N}$ be binomial random variables. Define

$$Y_n := \operatorname{var}(X_n)^{-1/2}\big(X_n - \mathbb{E}\,X_n\big) = \frac{X_n - np}{\sqrt{np(1-p)}}\,. \tag{A.1.2}$$

Then the sequence Y_n, $n \in \mathbb{N}$, does not converge to the standard normal distribution $\mathcal{N}(0,1)$ in the total variation metric: letting $A := \big\{k/\sqrt{np(1-p)}\;;\; k \in \mathbb{Z}, n \in \mathbb{N}\big\}$, we have $\mathbb{P}(Y_n \in A) = 1$ for all $n \in \mathbb{N}$, while $\mathcal{N}(0,1)\{A\} = 0$. However, by the classical Laplace central limit theorem, it converges in d_{K}.

A.2 The Wasserstein metric

As seen in the previous section, convergence of a sequence of points does not imply convergence of the underlying sequence of the Dirac measures in all metrics. Here, we define a metric for which this is true. However, this metric will not be defined on all probability measures.

Definition A.2.1. Let S be endowed with a metric d and let $\mathscr{S}$ be the underlying Borel σ-algebra. A probability measure $\mu \in \Pr(S, \mathscr{S})$ has *finite first absolute moment* with respect to the underlying metric d) if

$$\int d(x, y)\,\mu(\mathrm{d}y) < \infty \tag{A.2.1}$$

for some (all) $x \in S$. The space of Borel probability measures with finite first absolute moment with respect to a metric d will be denoted by $\Pr^{L^1}(S, d)$.

Remark A.2.2. For each probability measure $\mu \in \Pr^{L^1}(S, d)$ and each Lipschitz function $f\colon\ S \to \mathbb{R}$, we have $\langle |f|\,,\,\mu\rangle < \infty$.

Definition A.2.3. The *Wasserstein metric* (also known as Dudley, Fortet–Mourier or Kantorovich metric) on $\Pr^{L^1}(S, d)$ is defined by

$$d_{\mathrm{W}}(\mu, \nu) := \sup_{M_1(f) \le 1} \big|\langle f\,,\,\nu\rangle - \langle f\,,\,\mu\rangle\big|\,, \tag{A.2.2}$$

where

$$M_1(f) := \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)} \tag{A.2.3}$$

(notice that $M_1(f)$ is defined differently in (B.1.2), but the definitions coincide by Proposition B.1.9).

Proposition A.2.4. *The right hand side of* (A.2.2) *is finite for any two probability measures* $\mu, \nu \in \Pr^{L^1}(S,d)$.

Proof. First, we check finiteness. Taking f with $M_1(f) \leq 1$ and choosing arbitrary $z \in S$, observe that

$$\begin{aligned} \big|\langle f\,,\,\nu\rangle - \langle f\,,\,\mu\rangle\big| &= \left|\int_S\int_S \big(f(x)-f(y)\big)\,\mu(\mathrm{d}x)\,\nu(\mathrm{d}y)\right| \\ &\leq \int_S\int_S d(x,y)\,\mu(\mathrm{d}x)\,\nu(\mathrm{d}y) \\ &\leq \int_S\int_S \big(d(x,z)+d(y,z)\big)\,\mu(\mathrm{d}x)\,\nu(\mathrm{d}y) \\ &= \int_S d(x,z)\,\mu(\mathrm{d}x) + \int_S d(y,z)\,\nu(\mathrm{d}y) < \infty\,. \end{aligned} \tag{A.2.4}$$

From Proposition A.1.7, it follows that Lipschitz functions separate probability measures. Therefore, by Proposition A.1.1, d_{W} is indeed a metric. More precisely, the conditions of Proposition A.1.7 are fulfilled because the indicator of any closed set can be expressed as a limit of a uniformly bounded sequence of Lipschitz functions, while closed sets form a π-system generating the Borel σ-algebra $\mathscr{S}$. This completes the proof. □

There are several alternative definitions of the Wasserstein metric. A very important one, stated here as a theorem, is based on *couplings*.

Theorem A.2.5. *If S is separable, then for any two probability measures $\mu, \nu \in \Pr^{L^1}(S,d)$, we have*

$$d_{\mathrm{W}}(\mu,\nu) = \inf \mathbb{E}\big[d(X,Y)\big]\,, \tag{A.2.5}$$

where the minimum runs over all pairs of random variables X and Y defined on the same probability space, where X follows the distribution μ and Y follows the distribution ν.

Partial proof. For $X \sim \mu$ and $Y \sim \nu$ being defined on the same probability space and for $f\colon S \to \mathbb{R}$ with $M_1(f) \leq 1$, observe that

$$\big|\langle f\,,\,\mu\rangle - \langle f\,,\,\nu\rangle\big| = \big|\mathbb{E}\big[f(X)\big] - \mathbb{E}\big[f(Y)\big]\big| \leq \mathbb{E}\big|f(X)-f(Y)\big| \leq \mathbb{E}\big[d(X,Y)\big]\,.$$

Taking the supremum over all f in the left hand side and the infimum over all appropriate pairs (X,Y) in the right hand side, we find that $d_{\mathrm{W}}(\mu,\nu) \leq \inf \mathbb{E}\big[d(X,Y)\big]$. The proof of the opposite inequality is much more difficult and will be omitted here, but see Rachev [25]. □

Clearly, the total variation distance is stronger than the Kolmogorov distance. However, these two metrics are in general uncomparable to the Wasserstein distance, although in certain special cases, comparison is possible.

Example A.2.6. On $\mathbb{Z}$, we have $M_1(\mathbf{1}_A) \leq 1$ for all $A \subseteq \mathbb{Z}$. Therefore, $d_{\mathrm{TV}} \leq d_{\mathrm{W}}$, so that the Wasserstein metric is uniformly stronger than the total variation metric.

Example A.2.7. If S is bounded, the total variation metric is stronger than the Wasserstein metric – see Corollary A.5.9.

Example A.2.8. On an unbounded metric space with at least one accumulation point, the total variation and the Wasserstein metric are uncomparable: neither is stronger than the other. This is shown by the following two counterexamples: first, take an accumulation point x, so that there exist a sequence of points $x_1, x_2, \ldots$, which are all different from x and converge to x. Clearly, $d_{\mathrm{TV}}(\delta_{x_n}, \delta_x) = 1$ and $d_{\mathrm{W}}(\delta_{x_n}, \delta_x) = d(x_n, x)$. Therefore, the sequence $\delta_{x_1}, \delta_{x_2}, \ldots$ converges to δ_x in the Wasserstein metric, but it does not converge in the total variation metric. Next, choose an arbitrary point $y \in S$. Since S is unbounded, there exists a sequence $y_1, y_2, \ldots$, such that $d(y_n, y) \geq n$ for all n. Now consider the sequence of probability measures $\mu_1, \mu_2, \ldots$ defined as

$$\mu_n := \begin{pmatrix} y & y_n \\ 1 - \frac{1}{n} & \frac{1}{n} \end{pmatrix} = \left(1 - \frac{1}{n}\right) \delta_y + \frac{1}{n}\, \delta_{y_n}\,. \tag{A.2.6}$$

Since $d_{\mathrm{TV}}(\mu_n, \delta_y) = 1/n$, this sequence converges to the Dirac measure δ_y in the total variation metric. However, taking the test function $f(x) := d(x, x_0)$, we find that $\langle f\,,\, \mu_n \rangle \geq 1$ for all $n \geq 1$, while $\langle f\,,\, \delta_{x_0} \rangle = 0$, so that the sequence μ_n does not converge to δ_{x_0} in the Wasserstein metric.

Considering the same two examples on the real line, we find that the Kolmogorov and the Wasserstein metric are uncomparable, too.

However, it turns out that for some probability measures ν, any neighbourhood of ν in the Kolmogorov metric restricted to $\mathrm{Pr}^{L^1}(S, d)$ is also a neighbourhood of ν in the Wasserstein metric. Consequently, any sequence of probability measures in $\mathrm{Pr}^{L^1}(S, d)$ which converges to ν in the Wasserstein metric also converges to ν in the Kolmogorov metric.

Proposition A.2.9. *If $\nu \in \mathrm{Pr}^{L^1}(\mathbb{R})$ is a probability measure with density bounded from above by B, then we can estimate*

$$d_{\mathrm{K}}(\mu, \nu) \leq \sqrt{2B\, d_{\mathrm{W}}(\mu, \nu)} \tag{A.2.7}$$

for all $\mu \in \mathrm{Pr}^{L^1}(\mathbb{R})$.

Proof. For any $a \in \mathbb{R}$, define $f_a(x) := \mathbf{1}(x \leq a)$, so that $d_{\mathrm{K}}(\mu, \nu) = \sup_{a \in \mathbb{R}} \big| \langle f_a\,,\, \mu \rangle - \langle f_a\,,\, \nu \rangle \big|$. Next, for any $a < b$, define

$$f_{a,b}(x) := \begin{cases} 1 & ;\ x \leq a \\ \dfrac{b - x}{b - a} & ;\ a \leq x \leq b \\ 0 & ;\ x \geq b \end{cases}$$

and observe that $M_1(f_{a,b}) = 1/(b-a)$. Therefore, $\big|\langle f_{a,b}\,,\,\mu\rangle - \langle f_{a,b}\,,\,\nu\rangle\big| \le d_{\mathrm{W}}(\mu,\nu)/(b-a)$. Next, observe that

$$\langle f_{a,b}\,,\,\nu\rangle - \langle f_a\,,\,\nu\rangle = \int_a^b \frac{b-x}{b-a}\,\nu(\mathrm{d}x) \le B\int_a^b \frac{b-x}{b-a}\,\mathrm{d}x = \frac{B(b-a)}{2}\,,$$
$$\langle f_b\,,\,\nu\rangle - \langle f_{a,b}\,,\,\nu\rangle = \int_a^b \frac{x-a}{b-a}\,\nu(\mathrm{d}x) \le B\int_a^b \frac{x-a}{b-a}\,\mathrm{d}x = \frac{B(b-a)}{2}\,.$$

Therefore, for any $a \in \mathbb{R}$ and $\varepsilon > 0$,

$$\langle f_a\,,\,\mu\rangle - \langle f_a\,,\,\nu\rangle \le \langle f_{a,a+\varepsilon}\,,\,\mu\rangle - \langle f_{a,a+\varepsilon}\,,\,\nu\rangle + \frac{B\varepsilon}{2}\,,$$
$$\langle f_a\,,\,\nu\rangle - \langle f_a\,,\,\mu\rangle \le \langle f_{a-\varepsilon,a}\,,\,\nu\rangle - \langle f_{a-\varepsilon,a}\,,\,\mu\rangle + \frac{B\varepsilon}{2}\,.$$

Noting that the functions $f_{a,a+\varepsilon}$ and $f_{a-\varepsilon,a}$ have Lipschitz constant $1/\varepsilon$ and combining both estimates with the bound in terms of the Wasserstein distance, we find that

$$\big|\langle f_a\,,\,\mu\rangle - \langle f_a\,,\,\nu\rangle\big| \le \frac{d_{\mathrm{W}}(\mu,\nu)}{\varepsilon} + \frac{B\varepsilon}{2}\,.$$

Taking the supremum over all a, we obtain

$$d_{\mathrm{K}}(\mu,\nu) \le \frac{d_{\mathrm{W}}(\mu,\nu)}{\varepsilon} + \frac{B\varepsilon}{2}\,.$$

Optimization over ε completes the proof. □

A.3 More on the Kolmogorov metric

The Kolmogorov metric measures the distance between two probability measures in terms of the indicators of half-lines. However, we often need to consider more general test functions. Here, we show that we can take functions with bounded total variation on the whole real line. We refer to the results listed in Section B.2.

In view of definition B.2.1, define the total variation of a function f on the whole real line as

$$V(f) := V(f;\mathbb{R}) = \sup \sum_{i=1}^{n} \big|f(x_i) - f(x_{i-1})\big|\,,$$

where the supremum runs over all possible finite sequences $x_0 \le x_1 \le x_2 \le \cdots \le x_n$. A function f has *bounded variation* if its total variation is finite.

Proposition A.3.1. *For any two probability measures μ and ν on $\big(\mathbb{R}, \mathscr{B}(\mathbb{R})\big)$, we have*

$$d_{\mathrm{K}}(\mu,\nu) = \sup\big\{\big|\langle f\,,\,\mu\rangle - \langle f\,,\,\nu\rangle\big|\,;\, V(f) \le 1\big\}\,.$$

Remark A.3.2. If f has bounded variation, $\langle f\,,\,\mu\rangle$ exists for all Borel measures μ because f is bounded and Borel measurable.

Corollary A.3.3. *For any two probability measures μ and ν on $\big(\mathbb{R}, \mathscr{B}(\mathbb{R})\big)$ and any function f with bounded variation, we have*

$$\big|\langle f\,,\,\mu\rangle - \langle f\,,\,\nu\rangle\big| \le d_{\mathrm{K}}(\mu,\nu)\,V(f)\,.$$

Proof of Proposition A.3.1. Letting $\delta := d_{\mathrm{K}}(\mu,\nu)$, it suffices to prove that

$$\big|\langle f\,,\,\mu\rangle - \langle f\,,\,\nu\rangle\big| \le \delta\,V(f) \tag{A.3.1}$$

for all functions $f\colon \mathbb{R}\to\mathbb{R}$ with bounded variation. We shall do it step by step, knowing that (A.3.1) holds for functions of the form $f(x) = \mathbf{1}(x \le a)$.

Step 1: assume $f(x) = \mathbf{1}(x < a)$. Letting $f_n(x) = \mathbf{1}\big(x \le a - \frac{1}{n}\big)$, observe that the sequence f_n is monotone increasing and converges pointwise to f. The desired inequality (A.3.1) now follows from the monotone convergence theorem.

Step 2: assume that f is a *basic jump function*, i. e.,

$$f(x) = \begin{cases} 0 & x < a \\ \theta & x = a \\ 1 & x > a \end{cases}$$

for some $\theta \in [0,1]$. Noting that $V(f) = 1$ and $f(x) = 1 - (1-\theta)\,\mathbf{1}(x \le a) - \theta\,\mathbf{1}(x < a)$ and applying Step 1, we deduce that (A.3.1) is true for f.

Step 3: assume that f is a bounded monotone increasing *jump function*, i. e., of the form $f = \sum_{n=1}^{\infty} c_n f_n$, where $\sum_{n=1}^{\infty} c_n < \infty$. By part (3) of Proposition B.2.7, we have $V(f) = \sum_{n=1}^{\infty} c_n$. Clearly, $\langle f\,,\,\mu\rangle = \sum_{n=1}^{\infty}\langle f_n\,,\,\mu\rangle$ and $\langle f\,,\,\nu\rangle = \sum_{n=1}^{\infty}\langle f_n\,,\,\nu\rangle$. Therefore, (A.3.1) is true for f.

Step 4: assume that f is absolutely continuous, bounded and monotone increasing. In this case, $f(x) = \int_{-\infty}^{x} f'(a)\,\mathrm{d}a$ (see Theorem 6.4.2 of Heil [20]). By part (3) of Proposition B.2.7, we have $V(f) = \int_{-\infty}^{\infty} f'(a)\,\mathrm{d}a$. Rewriting the integral as $f(x) = \int_{-\infty}^{\infty} f'(a)\,\mathbf{1}(x \ge a)\,\mathrm{d}a$ and applying Fubini's theorem, we find that $\langle f\,,\,\mu\rangle = \int_{-\infty}^{\infty} \mu\big([a,\infty)\big)\,f'(a)\,\mathrm{d}a$ and $\langle f\,,\,\nu\rangle = \int_{-\infty}^{\infty} \nu\big([a,\infty)\big)\,f'(a)\,\mathrm{d}a$, so that (A.3.1) is true for f.

Step 5: assume that f is bounded and monotone increasing. By Lemma 1.6.31 (iii) of Tao [33], f can be decomposed as $f = g + h$, where g is a jump function g is absolutely continuous, and both functions are bounded and monotone increasing. Again, by part (3) of Proposition B.2.7, we have $V(f) = V(g) + V(h)$. Therefore, (A.3.1) follows from Steps 3 and 4.

Step 6: assume the general case where f has bounded variation. In this case, (A.3.1) follows from the previous step and the Jordan decomposition theorem (Theorem B.2.4). □

A.4 Weak topologies

In Section A.1, we introduced a construction of a metric on the set of probability measures based on a given class of test functions $\mathscr{F}$. This metric induces a topology, which will be

referred to as the *metric topology* with respect to $\mathscr{F}$. However, the class $\mathscr{F}$ introduces another very natural topology.

Definition A.4.1. Let the set $\mathscr{M}$ of probability measures and the class $\mathscr{F}$ of test functions be such that $\mu(|f|) < \infty$ for all $\mu \in \mathscr{M}$ and $f \in \mathscr{F}$. The *weak topology* on $\mathscr{M}$ with respect to $\mathscr{F}$ is the weakest topology, such that the functionals $\mu \mapsto \langle f\,,\,\mu\rangle$ are continuous for all $f \in \mathscr{F}$. In other words, this is the topology with a subbasis consisting of all sets of the form

$$G(f,U) := \{\nu \in \mathscr{M}\,;\,\langle f\,,\,\nu\rangle \in U\}\,, \tag{A.4.1}$$

where $f \in \mathscr{F}$ and U is an open set in $\mathbb{R}$.

Remark A.4.2. For each $\mu \in \mathscr{M}$, one can easily check that the sets

$$N(\mu; f_1,\ldots,f_n;\varepsilon_1,\ldots,\varepsilon_n) := \left\{\nu \in \mathscr{M}\,;\,\left|\langle f_i\,,\,\nu\rangle - \langle f_i\,,\,\mu\rangle\right| < \varepsilon_i, i = 1,\ldots,n\right\}, \tag{A.4.2}$$

where $f_1,\ldots,f_n \in \mathscr{F}$ and $\varepsilon_1,\ldots,\varepsilon_n > 0$, form a fundamental system of neighbourhoods of μ. *Therefore, in this topology, a sequence of probability measures μ_n converges to a probability measure μ if and only if the sequence $\langle f\,,\,\mu_n\rangle$ converges to $\langle f\,,\,\mu\rangle$ for each $f \in \mathscr{F}$.* We say that the sequence μ_n *weakly converges* to μ.

Comparing the topology introduced here with the topology from Section A.1, we obtain the following result.

Proposition A.4.3. *For given $\mathscr{M}$ and $\mathscr{F}$, the underlying weak topology is weaker than the underlying metric topology.*

Proof. Let $K(\mu,r)$ denote the open ball about μ with radius r, that is, $K(\mu,r) = \{\nu\,;\,d(\mu,\nu) < r\}$. It suffices to prove that each set $N(\mu; f_1,\ldots,f_n;\varepsilon_1,\ldots,\varepsilon_n)$ contains an open ball of the form $K(\mu,r)$. However, this is fulfilled by choosing $r := \min\{\varepsilon_1,\ldots,\varepsilon_n\}$. □

Remark A.4.4. Typically, the underlying metric topology is *strictly* stronger than the underlying weak topology, as illustrated in Example A.4.10 below.

Topologies have numerous important properties. The following assertion concerns two of them, which are satisfied by all metric topologies. The proof is left to the reader.

Proposition A.4.5.

(1) *The weak topology defined by the subbasis given in* (A.4.1) *is Hausdorff if and only if the space $\mathscr{F}$ of test functions separates the probability measures in $\mathscr{M}$.*

(2) *If $\mathscr{F}$ is countable, then the weak topology is first countable.*

□

Definition A.4.6. Let S be endowed with a topology and let $\mathscr{S}$ be the underlying Borel σ-algebra. The *usual weak topology* on $\Pr(S,\mathscr{S})$ is the one with respect to the class of all bounded continuous functions.

Proposition A.4.7. *If S is metrizable, then the class of all bounded continuous functions on S separates all probability measures.*

Proof. We apply Proposition A.1.7: it suffices to prove that there exists a π-sistem $\mathscr{P}$, which generates $\mathscr{S}$ and is such that for any set $A \in \mathscr{P}$, there exists a uniformly bounded sequence of continuous functions which converges pointwise to the indicator $\mathbf{1}_A$ of the set A. However, since $\mathscr{S}$ is the Borel σ-algebra, the family of all closed sets, which is a π-system, generates $\mathscr{S}$. Thus, for any closed set A, we need to construct a uniformly bounded sequence of continuous functions which converges pointwise to $\mathbf{1}_A$. If A is empty, we may take all of them to be zero. Otherwise, take a metric d on S and put

$$f_n(x) := \big(1 - n\, d(x, A)\big)_+ .$$

This completes the proof. □

Combining Propositions A.4.5 and A.4.7 leads to

Corollary A.4.8. *If S is metrizable, then the usual weak topology on $\Pr(S, \mathscr{S})$ is Hausdorff.* □

The Hausdorff property is all that we shall need from the properties of the weak topology on a space of probability measures. However, much more can be proved.

Theorem A.4.9. *If S is separable and metrizable, then $\Pr(S, \mathscr{S})$ with the weak topology is also metrizable.*

We shall omit the proof. For the case where S is *complete*, see, for example, Rogers and Williams [27], pp. 205–209. This proof is very indirect (it refers to the Banach–Alaoglu theorem). However, the result can also be proved directly, by considering the *Lévy–Prokhorov metric* on $\Pr(S, \mathscr{S})$ based on a metric d in S:

$$\rho(\mu, \nu) := \inf\{\varepsilon \geq 0 \,;\, \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ for all closed sets } A\}\,, \tag{A.4.3}$$

where $A^\varepsilon := \{x \in S \,;\, d(x, A) < \varepsilon\}$. Some rather involved calculation shows that this is really a metric which induces the weak topology. A substantial part of this is derived in Ethier and Kurtz [18] on pages 96–110, in fact all except for the fact that the weak topology is stronger than the topology induced by the Lévy–Prokhorov metric. This has to be verified directly – referring just to sequences does not suffice.

Now we turn to the example where the metric topology is strictly stronger than the weak topology with respect to the same class of test functions.

Example A.4.10. Let S be endowed with a topology, under which S is metrizable and is not discrete, and let $\mathscr{S}$ be the Borel σ-algebra. Let $\mathscr{F}$ be the class of all bounded continuous functions $S \to [0, 1]$. It is clear that the weak topology with respect to this class is just the usual weak topology. We shall show that it is strictly weaker than the underlying metric topology, noting that $d_{\mathscr{F}}$ is a metric because $\mathscr{F}$ separates all probability measures by Proposition A.4.7.

Since S is not discrete, there exists a point x, such that $\{x\}$ is not an open set. Since S is metrizable, there exists a sequence of points x_n different from x, which converges to x. Then it is clear that the sequence of the Dirac measures δ_{x_n} weakly converges to δ_x. On the other hand, it is clear that for each $n \in \mathbb{N}$, there exists a function $f \in \mathscr{F}$, such that $f(x) = 0$ and $f(x_n) = 1$, so that $d_{\mathscr{F}}(\delta_x, \delta_{x_n}) = 1$. As a result, the sequence does not converge in the metric topology.

On the real line, the usual weak topology is characterized by the following property.

Proposition A.4.11. *Let $\mu_1, \mu_2, \ldots,$ and μ be Borel probability measures on the real line with the underlying cumulative distribution functions $F_1, F_2, \ldots$ and F, that is, $F_n(a) = \mu_n\big((-\infty, a]\big)$ and $F(a) = \mu\big((-\infty, a]\big)$. Then the sequence $\mu_1, \mu_2, \ldots$ converges to μ in the usual weak topology if and only if the values $F_1(c), F_2(c), \ldots$ converge to $F(a)$ for each c where F is continuous, i. e., where $\mu(\{c\}) = 0$.*

Proof. Suppose first that the measures weakly converge. Similarly as in the proof of Proposition A.2.9, for any $a < b$, define

$$f_{a,b}(x) := \begin{cases} 1 & ; x \leq a \\ \dfrac{b-x}{b-a} & ; a \leq x \leq b \\ 0 & ; x \geq b \end{cases} .$$

Observe that $f_{a,b}(x) = \frac{1}{b-a} \int_a^b \mathbf{1}_{(-\infty,t]}(x)\,\mathrm{d}t$, so that

$$\langle f_{a,b}\,,\,\mu\rangle = \frac{1}{b-a}\int_a^b F(t)\,\mathrm{d}t\,.$$

Now take a point c where F is continuous and take $\varepsilon > 0$. There exists $\delta > 0$, such that $|F(x) - F(c)| < \varepsilon/2$ for all x with $|x - c| < \delta$. Consequently,

$$F(c) - \frac{\varepsilon}{2} < \langle f_{c-\delta,c}\,,\,\mu\rangle \leq F(c) \leq \langle f_{c,c+\delta}\,,\,\mu\rangle \leq F(c) + \frac{\varepsilon}{2}\,.$$

Since the measures weakly converge, there exists n_0, such that $\big|\langle f_{c,c+\delta}\,,\,\mu_n\rangle - \langle f_{c,c+\delta}\,,\,\mu\rangle\big| < \varepsilon/2$ and $\big|\langle f_{c-\delta,c}\,,\,\mu_n\rangle - \langle f_{c-\delta,c}\,,\,\mu\rangle\big| < \varepsilon/2$ for all $n \geq n_0$. Now estimate

$$F_n(c) - F(c) < \langle f_{c,c+\delta}\,,\,\mu_n\rangle - \langle f_{c,c+\delta}\,,\,\mu\rangle + \frac{\varepsilon}{2} < \varepsilon$$

and

$$F(c) - F_n(c) < \langle f_{c-\delta,c}\,,\,\mu\rangle + \frac{\varepsilon}{2} - \langle f_{c-\delta,c}\,,\,\mu_n\rangle < \varepsilon$$

This proves that the sequence $F_1(c), F_2(c), \ldots$ converges to $F(c)$.

Now we turn to the converse: suppose that the values $F_1(c), F_2(c), \ldots$ converge to $F(c)$ for each c where F is continuous. Equivalently, the sequence $\mu_1(\mathbf{1}_{(-\infty,c]}), \mu_2(\mathbf{1}_{(-\infty,c]}), \ldots$ converges to $\mu(\mathbf{1}_{(-\infty,c]})$. Consequently, for any function of the form

$$g = a_0 + \sum_{k=1}^{m} a_k\,\mathbf{1}_{(-\infty,c_k]}\,,$$

where F is continuous at all points $c_1, \ldots, c_m$ and where $a_0, a_1, \ldots, a_m$ are any real numbers, the sequence $\langle g\,,\,\mu_1\rangle, \langle g\,,\,\mu_2\rangle, \ldots$ converges to $\langle g\,,\,\mu\rangle$.

Now take a continuous function $f\colon\, \mathbb{R} \to \mathbb{R}$ with $|f(x)| \leq B$ for all $x \in \mathbb{R}$. Wa can assume that $B > 0$. Take $\varepsilon > 0$. Recalling that continuity points of F are everywhere dense, observe that there exist points $a < b$, where F is continuous and are such that $\mu\big(-\infty, a]\big) < \varepsilon/(15B)$ and $\mu\big([b, \infty)\big) < \varepsilon/(15B)$. Next, there exists n_0, such that $|F_n(a) - F(a)| < \varepsilon/(15B)$ and $|F_n(b) - F(b)| < \varepsilon/(15B)$ for all $n \geq n_0$. As a result, we have $\mu_n\big(-\infty, a]\big) < 2\varepsilon/(15B)$ and $\mu_n\big((b, \infty)\big) < 2\varepsilon/(15B)$.

Since f is uniformly continuous on $[a, b]$, there exists $\delta > 0$, such that $|f(x) - f(y)| < \varepsilon/15$ for all x, y with $|x - y| < \delta$. Next, there exist points $a = c_0 < c_1 < \cdots < c_m = b$, such that $c_k - c_{k-1} < \delta$ for all $k = 1, 2, \ldots, m$ and f is continuous at all c_k. Consider the function

$$g(x) = \begin{cases} f(c_1) & ;\ x \leq c_1 \\ f(c_2) & ;\ c_1 < x \leq c_2 \\ \cdots & \\ f(c_{m-1}) & ;\ c_{m-2} < x \leq c_{m-1} \\ f(c_m) & ;\ x > c_{m-1}\,. \end{cases}$$

Observe that $|g(x) - f(x)| < \varepsilon/15$ for all $a \leq x \leq b$ and $|g(x) - f(x)| < 2B$ for all other x. Therefore,

$$\begin{aligned} \big|\langle f\,,\,\mu\rangle - \langle g\,,\,\mu\rangle\big| &\leq \int_{(-\infty,a]} |g - f|\,\mathrm{d}\mu + \int_{(a,b]} |g - f|\,\mathrm{d}\mu + \int_{(b,\infty)} |g - f|\,\mathrm{d}\mu \\ &< 2B\,\frac{\varepsilon}{15B} + \frac{\varepsilon}{15} + 2B\,\frac{\varepsilon}{15B} = \frac{5\varepsilon}{15} \end{aligned}$$

and similarly,

$$\begin{aligned} \big|\langle f\,,\,\mu_n\rangle - \langle g\,,\,\mu_n\rangle\big| &\leq \int_{(-\infty,a]} |g - f|\,\mathrm{d}\mu_n + \int_{(a,b]} |g - f|\,\mathrm{d}\mu_n + \int_{(b,\infty)} |g - f|\,\mathrm{d}\mu_n \\ &< 2B\,\frac{2\varepsilon}{15B} + \frac{\varepsilon}{15} + 2B\,\frac{2\varepsilon}{15B} = \frac{9\varepsilon}{15} \end{aligned}$$

for all $n \geq n_0$. Finally, since

$$g = \sum_{k=1}^{m-1} \big(f(c_k) - f(c_{k+1})\big)\,\mathbf{1}_{(-\infty,c_k]} + f(c_m)\,,$$

the sequence $\langle g\,,\,\mu_1\rangle, \langle g\,,\,\mu_2\rangle, \ldots$ converges to $\langle g\,,\,\mu\rangle$. Therefore, there exists $n_1 \geq n_0$, such that $\big|\langle g\,,\,\mu_n\rangle - \langle g\,,\,\mu\rangle\big| < \varepsilon/15$ for all $n \geq n_1$ and we have

$$\big|\langle f\,,\,\mu_n\rangle - \langle f\,,\,\mu\rangle\big| \leq \big|\langle f\,,\,\mu_n\rangle - \langle g\,,\,\mu_n\rangle\big| + \big|\langle g\,,\,\mu_n\rangle - \langle g\,,\,\mu\rangle\big| + \big|\langle g\,,\,\mu\rangle - \langle f\,,\,\mu\rangle\big| < \varepsilon\,,$$

so that the sequence $\langle f\,,\,\mu_1\rangle, \langle f\,,\,\mu_2\rangle, \ldots$ indeed converges to $\langle f\,,\,\mu\rangle$. □

A.5 Change of class of test functions

We often consider different classes of test functions. In view of this, it is useful to compare the underlying metric and weak topologies. In this section, we shall consider three important relationships between the classes: inclusion, linear combinations and limits. The first case is immediate and is therefore stated without proof.

Proposition A.5.1. *Let $\mathscr{F}$ and $\mathscr{G}$ be classes of test functions with $\mathscr{F} \subseteq \mathscr{G}$. Assume that $\langle |g|\,,\,\mu\rangle < \infty$ for all $\mu \in \mathscr{M}$ and $g \in \mathscr{G}$.*

(1) *Let $d_{\mathscr{F}}$ and $d_{\mathscr{G}}$ be defined as in (A.1.1). If $d_{\mathscr{F}}$ is a metric on $\mathscr{M}$, so is $d_{\mathscr{G}}$ and we have $d_{\mathscr{F}} \leq d_{\mathscr{G}}$. Thus, the metric $d_{\mathscr{F}}$ is uniformly weaker than $d_{\mathscr{G}}$.*

(2) *The weak topology with respect to $\mathscr{F}$ is weaker than the weak topology with respect to $\mathscr{G}$.*

□

Proposition A.5.2. *Let $\mathscr{F}$ be a class of test functions. Assume that $\langle |f|\,,\,\mu\rangle < \infty$ for all $\mu \in \mathscr{M}$ and $f \in \mathscr{F}$.*

(1) *If*

$$\bar{\mathscr{F}} = \left\{ \alpha_0 + \sum_{i=1}^{n} \alpha_i f_i \,;\, f_i \in \mathscr{F},\, \sum_{i=1}^{n} |\alpha_i| \leq 1,\, n \in \mathbb{N} \right\}, \tag{A.5.1}$$

then the metric with respect to $\mathscr{F}$ agrees with the one with respect to $\bar{\mathscr{F}}$.

(2) *The weak topology with respect to $\mathscr{F}$ agrees with the weak topology with respect to* $\operatorname{span}\big(\{1\} \cup \mathscr{F}\big)$.

Corollary A.5.3. *If $\mathscr{F} \subseteq \mathscr{G} \subseteq \bar{\mathscr{F}}$, the metric with respect to $\mathscr{G}$ agrees with the metric with respect to $\mathscr{F}$. Similarly, if $\mathscr{F} \subseteq \mathscr{G} \subseteq \operatorname{span}\big(\{1\} \cup \mathscr{F}\big)$, the weak topology with respect to $\mathscr{G}$ coincides with the weak topology with respect to $\mathscr{F}$. In particular, the weak topology with respect to $\bar{\mathscr{F}}$ coincides with the weak topology with respect to $\mathscr{F}$.* □

Proof of Proposition A.5.2.

(1): Denote the underlying metrics by $d_{\mathscr{F}}$ and $d_{\bar{\mathscr{F}}}$ and take any probability measures $\mu, \nu \in \mathscr{M}$. Clearly, $d_{\mathscr{F}}(\mu,\nu) \leq d_{\bar{\mathscr{F}}}(\mu,\nu)$. However, we also have $\big|\langle f\,,\,\nu\rangle - \langle f\,,\,\mu\rangle\big| \leq d_{\mathscr{F}}(\mu,\nu)$ for all $f \in \bar{\mathscr{F}}$, so that $d_{\bar{\mathscr{F}}}(\mu,\nu) \leq d_{\mathscr{F}}(\mu,\nu)$.

(2): It suffices to prove that the weak topology with respect to $\operatorname{span}\big(\{1\} \cup \mathscr{F}\big)$ is weaker than the weak topology with respect to $\mathscr{F}$. In order to prove the latter, it suffice to prove that the set $G(f,U)$ defined as in (A.4.1) is open in the weak topology with respect to $\mathscr{F}$ for each $f \in \operatorname{span}\big(\{1\} \cup \mathscr{F}\big)$ and each open set $U \subseteq \mathbb{R}$. Take arbitrary $f = \alpha_0 + \sum_{i=1}^{n} \alpha_i f_i \in \operatorname{span}\big(\{1\} \cup \mathscr{F}\big)$, where $f_i \in \mathscr{F}$, and arbitrary $\mu \in G(f,U)$, that is, $\langle f\,,\,\mu\rangle = \alpha_0 + \sum_{i=1}^{n} \alpha_i \langle f_i\,,\,\mu\rangle \in U$. Since the map $(y_1, \ldots y_n) \mapsto \sum_{i=1}^{n} \alpha_i y_i$ is continuous,

there exist open sets $U_i \subseteq \mathbb{R}$, such that $\langle f_i\,,\,\mu\rangle \in U_i$ and $\alpha_0 + \sum_{i=1}^n \alpha_i y_i \in U$ for all $y_i \in U_i$. Therefore, $\mu \in \bigcap_{i=1}^n G(f_i, U_i) \subseteq G(f, U)$, implying that $G(f,U)$ is indeed open in the weak topology with respect to $\mathscr{F}$. This completes the proof. □

Proposition A.5.4. *Let $\mathscr{F}$ and $\mathscr{G}$ be classes of measurable test functions on $(S, \mathscr{S})$. Assume that $\langle |f|\,,\,\mu\rangle < \infty$ for all $\mu \in \mathscr{M}$ and $f \in \mathscr{F} \cup \mathscr{G}$.*

(1) *Let $d_{\mathscr{F}}$ and $d_{\mathscr{G}}$ denote the metrics with respect to $\mathscr{F}$ and $\mathscr{G}$, respectively, by (A.1.1). If for each $f \in \mathscr{F}$, there exists a sequence of test functions $f_n \in \mathscr{G}$, such that $\lim_{n\to\infty}\langle f_n\,,\,\mu\rangle = \langle f\,,\,\mu\rangle$ for all $\mu \in \mathscr{M}$, then $d_{\mathscr{F}} \leq d_{\mathscr{G}}$.*

(2) *If for each $f \in \mathscr{F}$ and each $\mu \in \mathscr{M}$, there exist sequences of functions $f_n^+, f_n^- \in \mathscr{G}$, such that $f_n^- \leq f \leq f_n^+$ and $\lim_{n\to\infty}\langle f_n^+\,,\,\mu\rangle = \lim_{n\to\infty}\langle f_n^-\,,\,\mu\rangle = \langle f\,,\,\mu\rangle$, then the weak topology on $\mathscr{M}$ with respect to $\mathscr{G}$ is stronger than the one with respect to $\mathscr{F}$.*

Corollary A.5.5. *Under the assumptions of the preceding lemma along with the assumption that $\mathscr{G} \subseteq \mathscr{F}$, the metrics as well as the weak topologies based on $\mathscr{F}$ and $\mathscr{G}$ agree.* □

Remarks.

(1) The conditions of part (1) are fulfilled if the sequence f_n converges to f pointwise and is either monotone or uniformly bounded.

(2) The conditions of part (2) are fulfilled if the sequences f_n^+ and f_n^- pointwise converge to f and one of the following two conditions is fulfilled: either both sequences are uniformly bounded or f_n^+ is monotonically decreasing, while f_n^- is monotonically increasing.

Proof of Proposition A.5.4.

(1): Immediate.

(2): For all $\mu \in \mathscr{M}$, $f \in \mathscr{F}$ and $\varepsilon > 0$, it suffices to construct a neighbourhood of μ in the weak topology with respect to $\mathscr{G}$, which is contained in the set $N(\mu; f; \varepsilon)$ defined as in (A.4.2). Taking $\varepsilon > 0$ and appropriate sequences f_n^+ and f_n^-, there exists n, such that $\langle f\,,\,f_n^+\rangle - \varepsilon/2 < \langle f\,,\,\mu\rangle < \langle f_n^-\,,\,\mu\rangle + \varepsilon/2$. Define

$$U := \left\{\nu \in \mathscr{M}\,;\, \langle f_n^+\,,\,\nu\rangle - \langle f_n^+\,,\,\mu\rangle < \frac{\varepsilon}{2},\ \langle f_n^-\,,\,\nu\rangle - \langle f_n^-\,,\,\mu\rangle < \frac{\varepsilon}{2}\right\}. \tag{A.5.2}$$

Clearly, $\mu \in U$ and U is $\mathscr{G}$-weakly open. Moreover,

$$\langle f\,,\,\nu\rangle \leq \langle f_n^+\,,\,\nu\rangle = \langle f_n^+\,,\,\nu\rangle - \langle f_n^+\,,\,\mu\rangle + \langle f_n^+\,,\,\mu\rangle < \langle f\,,\,\mu\rangle + \varepsilon \tag{A.5.3}$$

and

$$\langle f\,,\,\nu\rangle \geq \langle f_n^-\,,\,\nu\rangle = \langle f_n^-\,,\,\nu\rangle - \langle f_n^-\,,\,\mu\rangle + \langle f_n^-\,,\,\mu\rangle > \langle f\,,\,\mu\rangle - \varepsilon \tag{A.5.4}$$

for all $\nu \in U$. Therefore, $U \subseteq N(\mu; f; \varepsilon)$ and the proof is complete. □

As a simple example, consider the following characterization of the total variation metric.

Proposition A.5.6. *We have*

$$d_{\mathrm{TV}}(\mu,\nu) = \sup_{\substack{f \text{ is measurable} \\ 0\leq f\leq 1}} \big|\langle f\,,\,\nu\rangle - \langle f\,,\,\mu\rangle\big|\,. \tag{A.5.5}$$

Remark A.5.7. For the total variation metric, some authors take the following one instead:

$$d(\mu,\nu) = \sup_{\substack{f \text{ is measurable} \\ -1\leq f\leq 1}} \big|\langle f\,,\,\nu\rangle - \langle f\,,\,\mu\rangle\big| = 2\,d_{\mathrm{TV}}(\mu,\nu)\,. \tag{A.5.6}$$

Proof of Proposition A.5.6. By part (1) of Proposition A.5.2, the class of indicators can be replaced by the class of all measurable step functions from S to $[0,1]$. This is because each measurable $[0,1]$-valued step function can be represented as

$$f(x) = \begin{cases} y_0 & ;\, x\in H_0 \\ y_1 & ;\, x\in H_1 \\ \vdots & \quad\vdots \\ y_n & ;\, x\in H_n\,, \end{cases}$$

where $0 = y_0 \leq y_1 \leq \cdots \leq y_n \leq 1$ and where $H_0, H_1, \ldots$ are disjoint measurable sets, and this can be further expressed as

$$f = y_1\,\mathbf{1}_{H_1} + (y_2-y_1)\,\mathbf{1}_{H_1\cup H_2} + \cdots + (y_n - y_{n-1})\,\mathbf{1}_{H_1\cup H_2\cup\cdots\cup H_n}\,,$$

noting that $\sum_{k=1}^{n}(y_k - y_{k-1}) = y_n \leq 1$.

For each measurable $[0,1]$-valued function, there exists a monotone sequence of measurable $[0,1]$-valued step functions converging pointwise to it. Therefore, by part (1) of Proposition A.5.4, the class of all measurable $[0,1]$-valued step functions can be further replaced by the class of all measurable functions from S to $[0,1]$. This completes the proof. □

Corollary A.5.8. *The topology induced by the total variation metric is stronger than the usual weak topology.*

Corollary A.5.9. *If S is a bounded metric space, then the total variation metric on $\mathrm{Pr}(S,\mathscr{S})$ is stronger than the Wasserstein metric. More precisely, if D is the diameter of S, we have $d_{\mathrm{W}}(\mu,\nu) \leq D\,d_{\mathrm{TV}}(\mu,\nu)$.* □

Proposition A.5.10. *Let $\mathscr{F}$ be a class of test functions on $(S,\mathscr{S})$, such that $\langle |f|\,,\,\mu\rangle < \infty$ for all $\mu\in\mathscr{M}$. In addition, suppose that at least one of the following two conditions is fulfilled:*

(1) *For each open set $G\subseteq S$, there exists a sequence of functions $f_n\in\mathscr{F}$, $f_n\leq \mathbf{1}_G$, which is uniformly bounded from below and converges pointwise to $\mathbf{1}_G$.*

(2) *For each closed set $F\subseteq S$, there exists a sequence of functions $f_n\in\mathscr{F}$, $f_n\geq \mathbf{1}_F$, which is uniformly bounded from above and converges pointwise to $\mathbf{1}_F$.*

Then the weak topology with respect to $\mathscr{F}$ is stronger than the usual weak topology on $\mathscr{M}$.

Proof. First observe that, by part (2) of Proposition A.5.2, we may assume without loss of generality that $\mathscr{F}$ is a vector space and that it contains all constants. Under this assumption, $1 - f \in \mathscr{F}$ for all $f \in \mathscr{F}$, so that conditions (1) and (2) are equivalent. We may assume that both are fulfilled.

We shall use part (2) of Proposition A.5.4: for each bounded and continuous $f\colon S \to \mathbb{R}$ and each $\mu \in \mathscr{M}$, we shall construct sequences of functions $f_n^+, f_n^- \in \mathscr{F}$, such that $f_n^- \leq f \leq f_n^+$ and $\lim_{n\to\infty}\langle f_n^+ , \mu\rangle = \lim_{n\to\infty}\langle f_n^- , \mu\rangle = \langle f , \mu\rangle$. Without loss of generality, we can assume that f is $[0,1]$-valued. For any $t \in \mathbb{R}$, let:

$$F_t := \{x\,;\, f(x) \geq t\}\,, \qquad G_t := \{x\,;\, f(x) > t\} \tag{A.5.7}$$

One can easily check that

$$f - \frac{1}{n} \leq \frac{1}{n}\sum_{k=1}^{n-1} \mathbf{1}_{G_{k/n}} \leq f \leq \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_{F_{k/n}} \leq f + \frac{1}{n} \tag{A.5.8}$$

for all $n \in \mathbb{N}$. From assumption (1) and the dominated convergence theorem, it follows that there exist functions $f_{n,1}^-, \ldots, f_{n,n-1}^- \in \mathscr{F}$, such that $f_{n,k}^- \leq \mathbf{1}_{G_{k/n}}$ and $\langle f_{n,k}^- , \mu\rangle > \mu(G_{k/n}) - 1/n$ for all k. Similarly, from assumption (2), it follows that there exist functions $f_{n,0}^+, \ldots, f_{n,n-1}^+ \in \mathscr{F}$, such that $f_{n,k}^+ \geq \mathbf{1}_{F_{k/n}}$ and $\langle f_{n,k}^+ , \mu\rangle < \mu(F_{k/n}) + 1/n$ for all k. Put

$$f_n^+ := \sum_{k=0}^{n-1} f_{n,k}^+\,, \qquad f_n^- := \sum_{k=1}^{n-1} f_{n,k}^-\,. \tag{A.5.9}$$

Clearly, $f_n^- \leq f \leq f_n^+$. A short calculation shows that

$$\langle f , \mu\rangle - \frac{2}{n} \leq \langle f_n^- , \mu\rangle \leq \langle f , \mu\rangle \leq \langle f_n^+ , \mu\rangle \leq \langle f , \mu\rangle + \frac{2}{n}\,, \tag{A.5.10}$$

so that the sequences f_n^- and f_n^+ satisfy the necessary conditions and the proof is complete. □

Corollary A.5.11. *If S is endowed with a metric d, then the topology induced by the Wasserstein metric is stronger than the usual weak topology.*

Proof. First, by Proposition A.4.3, the topology induced by the Wasserstein metric is stronger than the weak topology with respect to the class of functions f with $M_1(f) \leq 1$. By part (2) of Proposition A.5.2, the latter coincides with the topology with respect to the class of all Lipschitz functions. Now observe that the latter class fulfills condition (2) of Proposition A.5.10: for a closed set F, the latter is satisfied with

$$f_n(x) := \big(1 - n\, d(x, F)\big)_+ \tag{A.5.11}$$

(under the convention that $d(x, \emptyset) = \infty$). This completes the proof. □

Corollary A.5.12. *The topology on* $\Pr\big(\mathbb{R}, \mathscr{B}(\mathbb{R})\big)$ *induced by the Kolmogorov metric is stronger than the usual weak topology.*

Proof. Again, by Proposition A.4.3, the topology induced by the Kolmogorov metric is stronger than the weak topology with respect to the class of indicators of all closed half-lines. By part (2) of Proposition A.5.2, the latter coincides with the topology with respect to the class of all linear combinations of indicators of closed half-lines. Now we show that this class fulfills condition (1) of Proposition A.5.10. To see this, observe that each open set G on the real line is a countably infinite union of disjoint intervals of form $(u, v]$. Thus, we may write

$$\mathbf{1}_G = \sum_{k=1}^{\infty} \mathbf{1}_{(u_k, v_k]} \ .$$

As the underlying partial sums are linear combinations of indicators of closed half-lines and are smaller than $\mathbf{1}_G$, condition (1) is indeed satisfied. This completes the proof. □

Appendix B

Some real analysis

B.1 Differentiation of absolutely continuous functions

Throughout this section, let I denote an interval on the real line.

The classical fundamental theorem of calculus says that a *continuously differentiable* function $f\colon I \to \mathbb{R}$ satisfies

$$f(b) - f(a) = \int_a^b f'(x)\,\mathrm{d}x \tag{B.1.1}$$

for all $a, b \in I$. This is also true for many functions which are only *almost everywhere* differentiable, i. e., $f(x) = |x|$.

Definition B.1.1. A measurable function $f'\colon T \to \mathbb{R}$ is an *almost-everywhere* derivative of a function $f\colon I \to \mathbb{R}$ if the set of points in I where f' is not the derivative of f has Lebesgue measure zero.

However, if f' is an almost-everywhere derivative of f, the fundamental theorem of calculus (B.1.1) need not be true. A prominent counter-example is the Cantor–Lebesgue function – see Theorem 5.1.2 of Heil [20].

The Cantor–Lebesgue function is continuous, but its continuity is not 'nice' enough. It turns out that the validity of (B.1.1) is equivalent to a stronger form of continuity.

Definition B.1.2. A function $f\colon [a,b] \to \mathbb{R}$ is *absolutely continuous* if for each $\varepsilon > 0$, there exists $\delta > 0$, such that for any (finite or infinite) family of non-overlapping intervals $[a_k, b_k] \subseteq [a,b]$ with $\sum_k (b_k - a_k) < \delta$, we have $\sum_k \big|f(b_k) - f(a_k)\big| < \varepsilon$.

Remark B.1.3. A restriction of an absolutely continuous function to a subinterval is also absolutely continuous.

Definition B.1.4. A function $f\colon I \to \mathbb{R}$ is absolutely continuous if all restrictions to f to closed subintervals of I are absolutely continuous.

Example B.1.5. Every Lipschitz function $f\colon I \to \mathbb{R}$ is absolutely continuous.

The following result states that absolute continuity is sufficient and necessary for the fundamental theorem of calculus to hold. For the proof, the reader is referred to Heil [20], Theorem 6.4.2.

Definition B.1.6. A function $f\colon I \to \mathbb{R}$ is *locally Lebesgue integrable* if all restrictions of f to closed subintervals of I are Lebesgue integrable. The space of all locally Lebesgue integrable functions on I will be denoted by $L^1_{\mathrm{loc}}(I)$.

Theorem B.1.7. *A function $f\colon I \to \mathbb{R}$ is absolutely continuous if and only if it is differentiable almost everywhere on I and some/any almost-everywhere derivative f' of f is in $L^1_{\mathrm{loc}}(I)$ (with respect to the Lebesgue measure) and satisfies (B.1.1) for all $a, b \in I$; of course, the integral in the right hand side of (B.1.1) is interpreted as the Lebesgue integral over the underlying interval or its negative value if $a > b$.*

Convention B.1.8. In the context of absolutely continuous functions f, f' will denote an almost-everywhere derivative of f unless specified otherwise.

In particular, all bounded measurable functions on I are in $L^1_{\mathrm{loc}}(I)$. For a function $f\colon I \to \mathbb{R}$, define

$$M_1(f) := \begin{cases} \operatorname{ess\,sup}_{x\in I} |f'(x)| & ;\ f \text{ is absolutely continuous} \\ \infty & ;\ \text{otherwise} \end{cases} \tag{B.1.2}$$

(clearly, the definition is independent of the version of f'). Next, for $r = 2, 3, 4, \ldots$, define

$$M_r(f) := \begin{cases} M_1\big(f^{(r-1)}\big) & ;\ f \text{ is } (r-1)\text{-times continuously differentiable on } I \\ \infty & ;\ \text{otherwise.} \end{cases} \tag{B.1.3}$$

Proposition B.1.9. *For any function $f\colon I \to \mathbb{R}$, we have*

$$M_1(f) = \sup_{x\neq y} \frac{|f(x) - f(y)|}{|x - y|}\,.$$

Sketch of proof. Denote $L(f) := \sup_{x\neq y} \frac{|f(x)-f(y)|}{|x-y|}$. First, we show that $M_1(f) \leq L(f)$. Without loss of generality, we may assume that $L(f) < \infty$. In this case, f is Lipschitz and therefore absolutely continuous. By Theorem B.1.7, there exists a function f', such that $f'(x)$ is the classical derivative of f at x for Lebesgue-almost all $x \in I$. However, for such x, we have $|f'(x)| \leq L(f)$. Therefore, $|f'(x)| \leq L(f)$ for Lebesgue-almost all x. As a result, $M_1(f) \leq L(f)$.

Now we prove that $L(f) \leq M_1(f)$. Similarly as before, we may assume that $M_1(f) < \infty$. Therefore, f is absolutely continuous. By Theorem B.1.7, we have $|f(b) - f(a)| = \left|\int_a^b f'(x)\,\mathrm{d}x\right| \leq |b-a| \operatorname{ess\,sup}_{x\in I} |f'(x)| = |b-a|\, M_1(f)$. Dividing by $|b-a|$ and taking supremum over a and b, we find that $L(f) \leq M_1(f)$, completing the proof. □

B.2 Functions with bounded variation

Like in the previous section, I will denote an interval on the real line throughout this section.

Definition B.2.1. The *total variation* of a function $f\colon I \to \mathbb{R}$ is defined as

$$V(f) := \sup \sum_{i=1}^{n} \big| f(x_i) - f(x_{i-1}) \big| ,$$

where the supremum runs over all possible finite sequences $x_0 \leq x_1 \leq x_2 \leq \cdots \leq x_n$ of points in I. The function f can be initially defined on a larger set than I. In this case, denote by $V(f; I)$ the total variation of the restriction of f to I, which will be also referred as the total variation of f on I.

A function f has *bounded variation* on I if its total variation on I is finite.

Remark B.2.2. Each function with bounded variation is bounded.

The proof of the following assertion is left to the reader as an exercise.

Proposition B.2.3. *If $f\colon I \to \mathbb{R}$ is monotone increasing or monotone decreasing, then*

$$V(f) = \sup f - \inf f .$$

Moreover, let $a_1, a_2, \ldots$ be a sequence of points in I chosen as follows: let a be the lower and b the upper endpoint of I. If $a \in I$, let $a_n = a$ for all n. Otherwise, let a_n converge to a. Similarly, choose a sequence $b_1, b_2, \ldots$, replacing a with b. Then we have

$$V(f) = \lim_{n\to\infty} \big| f(b_n) - f(a_n) \big| .$$

□

Theorem B.2.4 (Jordan decomposition theorem). *For each function $f\colon I \to \mathbb{R}$ with bounded variation, there exist monotone increasing functions g and h, such that $f = g - h$. Moreover, g and h can be chosen so that $V(f) = V(g) + V(h)$ and that g and h are left/right-continuous in all points where so is f.*

Sketch of proof. The existence of g and h is proved in Wheeden and Zygmund [34] for the case $I = [a, b]$, where $-\infty < a \leq b < \infty$: see Theorem 2.7 ibidem. From the construction of g and h, one can see that $V(f) = V(g) + V(h)$ and that g and h are left/right-continuous in all points where so is f (exercise). The extension to the general case is also left to the reader as an exercise. □

Corollary B.2.5. *Each function with bounded variation is Borel measurable.*

Sketch of proof. Observe first that each monotone increasing function has at most countably many discontinuities (see Theorem 2.8 of Wheeden and Zygmund [34]) and is therefore Borel measurable (exercise). The result now follows from the Jordan decomposition theorem. □

The following result if proved in Wheeden and Zygmund [34] – see Theorem 7.31 ibidem.

Proposition B.2.6. *If $f\colon I \to \mathbb{R}$ is absolutely continuous, then $V(f; I) = \int_I |f'(x)| \, \mathrm{d}x$.* □

Proposition B.2.7.

(1) *For any two functions g and $h\colon I \to \mathbb{R}$, we have $V(g+h) \le V(g) + V(h)$.*

(2) *Consider a measurable space $(S, \mathcal{S})$, a signed measure κ on S and a measurable function $F\colon I \times S \to \mathbb{R}$, such that the function $y \mapsto F(x,y)$ is an element of $L^1(|\kappa|)$ for all $x \in I$. Letting $f_y(x) := F(x,y)$ and $f(x) := \int_S F(x,y)\,\kappa(\mathrm{d}y)$, we have*

$$V(f) \le \int_S V(f_y)\,|\kappa|(\mathrm{d}y)\,. \tag{B.2.1}$$

(3) *If all functions f_y are monotone increasing and κ is a positive measure, then (B.2.1) holds with equality.*

Proof. It suffices to prove the second part. Taking $x_0 \le x_1 \le \cdots \le x_n$, observe that

$$\begin{aligned}\sum_{i=1}^n \big|f(x_i) - f(x_{i-1})\big| &= \sum_{i=1}^n \left|\int_S \big(F(x_i,y) - F(x_{i-1},y)\big)\,\kappa(\mathrm{d}y)\right| \\ &\le \int_S \sum_{i=1}^n \big|F(x_i,y) - F(x_{i-1},y)\big|\,|\kappa|(\mathrm{d}y) \\ &\le \int_S V(f_y)\,|\kappa(\mathrm{d}y)|\,.\end{aligned}$$

Taking the supremum, we obtain (B.2.1). Finally, if the functions $x \mapsto F(x,y)$ are monotone increasing and κ is a positive measure, take sequences a_n and b_n as in Proposition B.2.3, where, in addition, a_n is monotone decreasing and b_n is monotone increasing. Now observe that

$$\begin{aligned}V(f) &= \lim_{n\to\infty} \big(f(b_n) - f(a_n)\big) \\ &= \lim_{n\to\infty} \int_S \big(F(b_n,y) - F(a_n,y)\big)\,\kappa(\mathrm{d}y) \\ &= \int_S \lim_{n\to\infty} \big(F(b_n,y) - F(a_n,y)\big)\,\kappa(\mathrm{d}y) \\ &= \int_S V(f_y)\,\kappa(\mathrm{d}y)\end{aligned}$$

by the monotone convergence theorem. This completes the proof. □

B.3 The Riemann–Stieltjes integral

The results in this section are mostly just listed. For the proofs, the reader is referred to Chapter 2 and partly Chapter 7 of Wheeden and Zygmund [34], where the results may not be proved in the whole generality. However, the extension is not difficult and is left to the reader as an exercise.

Definition B.3.1. Let $-\infty < a \leq b < \infty$. Take functions $f, g\colon [a,b] \to \mathbb{R}$. The *Riemann–Stieltjes integral* of f with respect to g is the limit of the Riemann–Stieltjes sums

$$\sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1})\,,$$

where $a = x_0 \leq \xi_1 \leq x_1 \leq \xi_2 \leq \cdots \leq x_{n-1} \leq \xi_n \leq x_n = b$. More precisely, a number $J \in \mathbb{R}$ is the Riemann–Stieltjes integral of f with respect to g if for each $\varepsilon > 0$, there exists $\delta > 0$, such that

$$\left|\sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1}) - J\right| < \varepsilon$$

for all $x_0, \ldots, x_n$ and $\xi_1, \ldots, \xi_n$ being as above and such that $x_k - x_{k-1} < \delta$ for all $k = 1, 2, \ldots, n$. Write

$$J = \int_a^b f \,\mathrm{d}g\,.$$

The function f is called *integrand*, while the function g is called *integrator*.

Remark B.3.2. The Riemann–Stieltjes integral does not always exist. However, if it exists, it is unique.

Definition B.3.3. An *improper Riemann–Stieltjes integral* of f with respect to g is a limit of Riemann–Stieltjes integrals in the sense of Definition B.3.1 as one or both endpoints approach their limits – the left endpoint from the right and the right endpoint from the left. Thus, the integral can be improper at one or both endpoints. When the integral is improper at both endpoints, the limit of the integral must exist and be the same regardless how the endpoints approach their limits.

Denoting the limits of the endpoints by a and b (which can also be infinite), we also denote the underlying improper integral by $\int_a^b f \,\mathrm{d}g$, that is,

$$\int_a^b f \,\mathrm{d}g = \lim_{a' \downarrow a} \int_{a'}^b f \,\mathrm{d}g, \quad \lim_{b' \uparrow b} \int_a^{b'} f \,\mathrm{d}g \quad \text{or} \quad \lim_{\substack{a' \downarrow a \\ b' \uparrow b}} \int_{a'}^{b'} f \,\mathrm{d}g\,.$$

Remark B.3.4. When an endpoint is infinite, it is clear that the integral should be improper at that endpoint. However, when the endpoint is finite, we need to be careful as the proper and the underlying improper integral can both exist, but they can be different.

Unless specified otherwise, we the Riemann–Stieltjes integral will be consider proper at a certain endpoint if that endpoint is finite and both the integrand and the integrator are defined there.

Proposition B.3.5 ([34], Theorem 2.16). *The following is true for proper as well as improper Lebesgue–Stieltjes integrals, provided that all improper integrals in the same formula are considered in the same sense (e. g., all improper at the left endpoint):*

(1) *If $c \in \mathbb{R}$ and $\int_a^b f \,\mathrm{d}g$ exists, so do $\int_a^b (cf) \,\mathrm{d}g$ and $\int_a^b f \,\mathrm{d}(cg)$ and we have*

$$\int_a^b (cf) \,\mathrm{d}g = \int_a^b f \,\mathrm{d}(cg) = c \int_a^b f \,\mathrm{d}g\,.$$

(2) *If* $\int_a^b f_1 \,\mathrm{d}g$ *and* $\int_a^b f_2 \,\mathrm{d}g$ *exist, so does* $\int_a^b (f_1 + f_2) \,\mathrm{d}g$ *and we have*

$$\int_a^b (f_1 + f_2) \,\mathrm{d}g = \int_a^b f_1 \,\mathrm{d}g + \int_a^b f_2 \,\mathrm{d}g \,.$$

(3) *If* $\int_a^b f \,\mathrm{d}g_1$ *and* $\int_a^b f \,\mathrm{d}g_2$ *exist, so does* $\int_a^b f \,\mathrm{d}(g_1 + g_2)$ *and we have*

$$\int_a^b f \,\mathrm{d}(g_1 + g_2) = \int_a^b f \,\mathrm{d}g_1 + \int_a^b f \,\mathrm{d}g_2 \,.$$

□

Proposition B.3.6 ([34], Theorem 2.21). *If the proper integral* $\int_a^b f \,\mathrm{d}g$ *exists, then so does* $\int_a^b g \,\mathrm{d}f$ *and we have*

$$\int_a^b f \,\mathrm{d}g = f(b)\, g(b) - f(a)\, g(a) - \int_a^b g \,\mathrm{d}f \,.$$

For improper integrals, the statement remains true with $f(a)\, g(a)$ *and/or* $f(b)\, g(b)$ *replaced by the corresponding limits, which are assumed to exist. Thus, if the integral* $\int_a^b f \,\mathrm{d}g$ *exists in the improper sense at the right endpoint and* $\lim_{b' \uparrow b} f(b')\, g(b')$ *exists, then the integral* $\int_a^b g \,\mathrm{d}f$ *also exists in the improper sense at the right endpoint and we have*

$$\int_a^b f \,\mathrm{d}g = \lim_{b' \uparrow b} f(b')\, g(b') - f(a)\, g(a) - \int_a^b g \,\mathrm{d}f \,.$$

□

Proposition B.3.7 ([34], Chapter 2, Exercise 16). *The Riemann–Stieltjes integral* $\int_a^b f \,\mathrm{d}g$ *exists if* f *is continuous and* g *is of bounded variation on* $[a, b]$. □

Proposition B.3.8. *If* f *has bounded variation and* g *is absolutely continuous on* $[a, b]$*, then*

$$\int_a^b f \,\mathrm{d}g = \int_a^b f(x)\, g'(x) \,\mathrm{d}x \,. \tag{B.3.1}$$

Replacing the closed interval $[a, b]$ *with an open or half-open one (in this case, the endpoint which is not included can be infinite), the result remains true in the improper sense: if* f *has bounded total variation on all closed subintervals,* g *is absolutely continuous and the right hand side exists as an improper integral, so does the left hand side and they agree.* □

Proof. We shall only prove the result for closed intervals; the extension to open and semi-open intervals is left to the reader as an exercise.

First, we show that both sides of (B.3.1) exist. The right hand side exists because f is bounded (Remark B.2.2) and $g' \in L^1([a, b])$ (Theorem B.1.7). Next, since g is continuous and f has bounded variation, $\int_a^b g \,\mathrm{d}f$ exists by Proposition B.3.7. The integral $\int_a^b f \,\mathrm{d}g$ then exists by Proposition B.3.6.

To prove that both sides in (B.3.8) agree, take an array of points $a = x_0^{(n)} \le x_1^{(n)} \le \cdots \le x_n^{(n)} = b$, $n \in \mathbb{N}$, such that $\lim_{n\to\infty} \max_{1\le k\le n}(x_k - x_{k-1}) = 0$, e. g., $x_k^{(n)} = a + \frac{k}{n}(b-a)$. Write

$$\int_a^b f(x)\, g(x)\, \mathrm{d}x = A_n + B_n\,,$$

where

$$A_n := \sum_{k=1}^n f(x_k) \int_{x_{k-1}}^{x_k} g'(x)\, \mathrm{d}x = \sum_{k=1}^n f(x_k)\big(g(x_k) - g(x_{k-1})\big)\,,$$
$$B_n := \sum_{k=1}^n \int_{x_{k-1}}^{x_k} \big(f(x) - f(x_k)\big)\, g'(x)\, \mathrm{d}x = \int_a^b \delta_n(x)\, g'(x)\, \mathrm{d}x$$

and $\delta_n(x) := f(x) - f(x_k)$ for $x_{k-1} < x \le x_k$ and $\delta_n(a) := 0$. Clearly, $\lim_{n\to\infty} A_n = \int_a^b f\, \mathrm{d}g$. Next, since f has bounded total variation, it has at most countably many discontinuities. However, if f is continuous at x, we have $\lim_{n\to\infty} \delta_n(x) = 0$. Thus, the latter holds for almost all $x \in [a,b]$. Since f is bounded and $g' \in L^1([a,b])$, we can apply the dominated convergence theorem to deduce that $\lim_{n\to\infty} B_n = 0$, completing the proof. □

B.4 Signed measures

Definition B.4.1. A *signed measure* on a measurable space $(S, \mathscr{S})$ is a function $\mu\colon \mathscr{S} \to \mathbb{R}$, such that for any sequence $A_1, A_2, \ldots$ of pairwise disjoint sets, we have

$$\mu\Big(\bigcup_{k=1}^\infty A_k\Big) = \sum_{k=1}^\infty \mu(A_k)\,. \tag{B.4.1}$$

Remark B.4.2. Here, we only deal with finite signed measures. In general, it is also possible to consider signed measures with possible infinite values.

Remark B.4.3. As the sum in the right hand side of (B.4.1) is independent of the order of the summands, the series must converge absolutely.

Definition B.4.4. The *positive* and *negative part* of a signed measure μ are set functions μ^+ and μ^- defined by

$$\mu^+(A) := \sup\big\{\mu(B)\,;\, B \subseteq A, B \in \mathscr{S}\big\}\,, \qquad \mu^-(A) := \sup\big\{-\mu(B)\,;\, B \subseteq A, B \in \mathscr{S}\big\}\,. \tag{B.4.2}$$

The following result can be deduced from Corollary 4.1.6, Proposition 4.1.7 and Exercise 6 of Section 4.1 in Cohn [13]:

Proposition B.4.5. *The functions μ^+ and μ^- defined as above are finite positive measures and we have $\mu = \mu^+ - \mu^-$.* □

The representation $\mu = \mu^+ - \mu^-$ is called the *Jordan decomposition* of μ.

Definition B.4.6. The *variation* of a signed measure μ defined on $(S, \mathscr{S})$ is the positive measure $|\mu| := \mu^+ + \mu^-$. The *total variation* of the μ is defined as $\|\mu\| := |\mu|(S)$.

Definition B.4.7. For a function $f \in L^1(|\mu|)$, define the *Lebesgue integral* of f with respect to μ as

$$\int f \,\mathrm{d}\mu := \int f \,\mathrm{d}\mu_+ - \int f \,\mathrm{d}\mu^- ,$$

where $\mu = \mu^+ - \mu^-$ is called the Jordan decomposition of μ.

Remark B.4.8. We have $\left|\int f \,\mathrm{d}\mu\right| \le \int |f| \,\mathrm{d}|\mu|$.

Definition B.4.9. Let $g\colon \mathbb{R} \to \mathbb{R}$ be a right-continuous function with bounded variation. The *Lebesgue–Stieltjes measure* associated to g is the signed measure Λ_g determined by

$$\Lambda_g\big((a,b]\big) = g(b) - g(a) \tag{B.4.3}$$

for all $a \le b$.

Remark B.4.10. By Dynkin's π–λ theorem (see, e. g., Theorem 1.6.2 of Cohn [13]), (B.4.3) determines Λ_g uniquely.

Proposition B.4.11. *Let g be a right-continuous function g with bounded variation.*

(1) *The Lebesgue–Stieltjes measure Λ_g exists and we have $\|\Lambda_g\| = V(g)$.*

(2) *For $a \le b$ and a continuous function $f\colon [a,b] \to \mathbb{R}$, the Riemann–Stieltjes integral $\int_a^b f \,\mathrm{d}g$ exists and we have*

$$\int_{(a,b]} f \,\mathrm{d}\Lambda_g = \int_a^b f \,\mathrm{d}g .$$

Proof. By Theorem B.2.4, we may write $g = h - k$, where g and h are monotone increasing functions which are left/right-continuous at each point where so is f. Moreover, $V(g) = V(h) + V(k)$. Thus, h and k are right-continuous. By Theorem 11.10 of Wheeden and Zygmund [34], there exist Lebesgue–Stieltjes measures Λ_h and Λ_k, which are finite positive measures. From (B.4.3) and Dynkin's π–λ theorem, it follows that $\Lambda_g = \Lambda_h - \Lambda_k$.

By Proposition B.2.3, we have $V(h) = \lim_{n\to\infty} h(n) - h(-n) = \lim_{n\to\infty} \Lambda_h\big((-n,n]\big) = \Lambda_h(\mathbb{R})$ and similarly $V(k) = \Lambda_k(\mathbb{R})$. For each Borel set $B \subseteq \mathbb{R}$, we have $\Lambda_g(B) \le \Lambda_h(B)$ and $-\Lambda_g(B) \le \Lambda_k(B)$. Taking the supremum over B and recalling (B.4.2), we obtain $\Lambda_g^+(\mathbb{R}) \le \Lambda_h(\mathbb{R}) = V(h)$ and $\Lambda_g^-(\mathbb{R}) \le \Lambda_k(\mathbb{R}) = V(k)$. Therefore, $\|\Lambda_g\| = \Lambda_g^+(\mathbb{R}) + \Lambda_g^-(\mathbb{R}) \le V(h) + V(k) = V(g)$.

For each $\varepsilon > 0$, there exists a sequence $x_0 \le x_1 \le \cdots \le x_n$, such that $\sum_{k=1}^n \big|g(x_k) - g(x_{k-1})\big| > V(g) - \varepsilon$. Now observe that $\sum_{k=1}^n \big|\Lambda_g\big((x_{k-1}, x_k]\big)\big| \le \sum_{k=1}^n \big|\Lambda_g^+\big((x_{k-1}, x_k]\big) - \Lambda_g^-\big((x_{k-1}, x_k]\big)\big| \le \sum_{k=1}^n \big(\Lambda_g^+\big((x_{k-1}, x_k]\big) + \Lambda_g^-\big((x_{k-1}, x_k]\big)\big) = \Lambda_g^+\big((x_0, x_n]\big) + \Lambda_g^-\big((x_0, x_n]\big) \le \Lambda_g^+(\mathbb{R}) + \Lambda_g^-(\mathbb{R}) = \|\Lambda_g\|$. Therefore, $\|\Lambda_g\| > V(g) - \varepsilon$ for all $\varepsilon > 0$. As a result, we have $\|\Lambda_g\| \ge V(g)$. Combining with the preceding paragraph, we find that $\|\Lambda_g\| = V(g)$, proving part (1)

By Proposition B.3.7, the Riemann–Stieltjes integrals $\int f\,\mathrm{d}h$ and $\int f\,\mathrm{d}k$ both exist. By Proposition B.3.5, so does $\int f\,\mathrm{d}g$ and we have $\int f\,\mathrm{d}g = \int f\,\mathrm{d}h - \int f\,\mathrm{d}k$. By Theorem 11.11 of Wheeden and Zygmund [34], we have $\int_{(a,b]} f\,\mathrm{d}\Lambda_h = \int_a^b f\,\mathrm{d}h$ and $\int_{(a,b]} f\,\mathrm{d}\Lambda_k = \int_a^b f\,\mathrm{d}k$. Since $\Lambda_g = \Lambda_h - \Lambda_k$, we conclude that $\int_{(a,b]} f\,\mathrm{d}\Lambda_g = \int_a^b f\,\mathrm{d}g$, proving part (2). □

Appendix C

On the Mills ratio

C.1 Basic properties

Definition C.1.1. Let ϕ denote the standard Gaussian density on $\mathbb{R}$, that is,

$$\phi(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} \tag{C.1.1}$$

and let Φ denote its cumulative distribution function:

$$\Phi(w) := \int_{-\infty}^{w} \phi(x)\,\mathrm{d}x\,. \tag{C.1.2}$$

The *Mills ratio* is the ratio between these two functions as follows:

$$\psi(w) := \frac{\Phi(w)}{\phi(w)} = e^{\frac{1}{2}w^2} \int_{-\infty}^{w} e^{-\frac{1}{2}x^2}\,\mathrm{d}x\,. \tag{C.1.3}$$

Remark C.1.2. In the literature, the Mills ratio is often defined slightly differently, as the function $w \mapsto \psi(-w)$ (in our notation). The benefit of that definition is that it gives a 'tame' function for positive w. However, Definition (C.1.3) also has its benefits.

The Mills ratio is important because it can serve to express solutions to the Stein equation. As one can easily check, the Mills ratio solves the equation

$$\psi'(w) = w\,\psi(w) + 1\,, \tag{C.1.4}$$

which is a version of the Stein equation (3.2.1). Repeated differentiation gives further derivatives, some of which are listed below:

$$\psi''(w) = (w^2+1)\,\psi(w) + w\,, \tag{C.1.5}$$
$$\psi'''(w) = (w^3+3w)\,\psi(w) + w^2 + 2\,, \tag{C.1.6}$$
$$\psi^{(4)}(w) = (w^4+6w^2+3)\,\psi(w) + w^3 + 5w\,. \tag{C.1.7}$$

The derivatives of the Mills ratio also satisfy a recurrent formula as stated below:

Proposition C.1.3. *For all $r \in \mathbb{N}$ and all $w \in \mathbb{R}$, we have*

$$\psi^{(r+1)}(w) = w\,\psi^{(r)}(w) + r\,\psi^{(r-1)}(w)\,. \tag{C.1.8}$$

Proof. For $r = 1$, the identity is immediate from (C.1.4) and (C.1.5). The induction step from r to $r+1$ can be performed by differentiating (C.1.8), leading to

$$\psi^{(r+2)}(w) = w\,\psi^{(r+1)}(w) + (r+1)\,\psi^{(r)}(w)\,, \tag{C.1.9}$$

which is exactly (C.1.8) with $r+1$ in place of r. □

Corollary C.1.4. *The r-th derivative of the Mills ratio can be expressed as*

$$\psi^{(r)}(w) = P_r(w)\,\psi(w) + Q_r(w)\,,$$

where P_r is a polynomial of degree r and Q_r is a polynomial of degree $r-1$. □

Proposition C.1.5. *The Mills ratio and all its derivatives are strictly positive functions, i. e., $\psi^{(r)}(w) > 0$ for all $r = 0, 1, 2, \ldots$ and all $w \in \mathbb{R}$.*

Proof. Introducing a new variable $t := w - x$ into the integral in (C.1.3), we obtain

$$\psi(w) = e^{\frac{1}{2}w^2} \int_0^\infty e^{-\frac{1}{2}(w-t)^2}\,\mathrm{d}t = \int_0^\infty e^{tw-\frac{1}{2}t^2}\,\mathrm{d}t\,.$$

Repeated differentiation under the integral sign gives

$$\psi^{(r)}(w) = \int_0^\infty t^r e^{tw-\frac{1}{2}t^2}\,\mathrm{d}t > 0 \tag{C.1.10}$$

(one can easily check that this works because all these integrals converge uniformly and absolutely). □

Combining (C.1.4) and (C.1.5) with Proposition C.1.5 (which, among others, implies that ψ is strictly increasing), we obtain the following upper and lower bound on ψ for negative arguments: for all $w > 0$, we have

$$\frac{w}{w^2+1} < \psi(-w) < \min\left\{\sqrt{\frac{\pi}{2}}\,,\ \frac{1}{w}\right\}.$$

Moreover, derivatives of ψ can be bounded similarly.

Proposition C.1.6. *For all $r \in \mathbb{N}_0$ and all $w > 0$, we have*

$$\psi^{(r)}(-w) < \frac{r!}{w^{r+1}}\,. \tag{C.1.11}$$

Proof. From (C.1.10), we obtain

$$\psi^{(r)}(-w) = \int_0^\infty t^r e^{-tw-\frac{1}{2}t^2}\,\mathrm{d}t < \int_0^\infty t^w e^{-tw}\,\mathrm{d}t = \frac{r!}{w^{r+1}}\,, \tag{C.1.12}$$

completing the proof. □

C.2 Repeated integrals of the Gaussian density

There is a strong relationship between the derivatives of the Mills ratio and the repeated integrals of the standard Gaussian density ϕ. Inductively, define

$$\Phi_0(w) := \phi(w)\,, \qquad \Phi_{r+1}(w) := \int_{-\infty}^{w} \Phi_r(x)\,\mathrm{d}x\,. \tag{C.2.1}$$

Proposition C.2.1. *The functions Φ_r are all well defined (i. e., all integrals in (C.2.1) converge and we have*

$$\Phi_{r+1}(w) = \frac{1}{r!}\,\phi(w)\,\psi^{(r)}(w) \tag{C.2.2}$$

for all $r \in \mathbb{N}_0$.

Proof. For $r = 0$, (C.2.2) is immediate from the definition of the Mills ratio. Now make the induction step from $r-1$ to r. Thus, assume that

$$\Phi_r(w) = \frac{1}{(r-1)!}\,\phi(w)\,\psi^{(r-1)}(w)\,. \tag{C.2.3}$$

First, it follows from (C.1.12) that the function Φ_r is integrable on all intervals $(-\infty, w]$ and that (C.2.2) holds in the limit as $w \to -\infty$. Therefore, it suffices to prove the *derivative* of (C.2.2). By the induction hypothesis (C.2.3), we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}w}\left[\Phi_{r+1}(w) - \frac{1}{r!}\,\phi(w)\,\psi^{(r)}(w)\right] &= \Phi_r(w) - \frac{1}{r!}\,\phi'(w)\,\psi^{(r)}(w) - \frac{1}{r!}\,\phi(w)\,\psi^{(r+1)}(w)\\ &= \frac{1}{r!}\,\phi(w)\Big[r\,\psi^{(r-1)}(w) + w\,\psi^{(r)}(w) - \psi^{(r+1)}(w)\Big]\,.\end{aligned} \tag{C.2.4}$$

However, by the recurrent formula (C.1.8), the preceding expression equals zero. This completes the proof. □

Corollary C.2.2. *For each $r \in \mathbb{N}$, there exists C_r, such that*

$$\Phi^{(r)}(-w) \le C_r\,e^{-\frac{1}{2}w^2}$$

for all $w \ge 0$.

Proof. By (C.2.2) and since $\psi^{(r)}$ is increasing by Proposition C.1.5, we can estimate $\Phi^{(r)} = \frac{1}{(r-1)!}\,\phi(w)\,\psi^{(r-1)}(-w) \le \frac{1}{(r-1)!}\,\phi(w)\,\psi^{(r-1)}(0) = \frac{\psi^{(r-1)}(0)}{(r-1)!\sqrt{2\pi}}\,e^{-\frac{1}{2}w^2}$. □

Proposition C.2.3. *For all $r \in \mathbb{N}_0$ and all $w \in \mathbb{R}$, we have*

$$\Phi_r(-w)\,\psi^{(r)}(w) + \Phi_r(w)\,\psi^{(r)}(-w) = 1\,. \tag{C.2.5}$$

Proof. For $r = 0$, (C.2.5) reduces to the obvious identity $\Phi(w) + \Phi(-w) = 1$. For $r \in \mathbb{N}$, we apply Proposition C.2.1, which reduces (C.2.5) to

$$J_r(w) := \phi(w)\Big[\psi^{(r-1)}(-w)\,\psi^{(r)}(w) + \psi^{(r-1)}(w)\,\psi^{(r)}(-w)\Big] = (r-1)!\,, \tag{C.2.6}$$

and use induction over r. For $r = 1$, (C.1.4) implies

$$\begin{aligned} J_1(w) &= \phi(w)\Big[\psi(-w)\,\psi'(w) + \psi(w)\,\psi'(-w)\Big] = \\ &= \phi(w)\Big[\psi(-w)\big(w\,\psi(w)+1\big) + \psi(w)\big(-w\,\psi(-w)+1\big)\Big] = \\ &= \phi(w)\big(\psi(-w)+\psi(w)\big) = \\ &= 1\,. \end{aligned} \tag{C.2.7}$$

Now perform the induction step from r to $r+1$. This time, application of the recurrent formula (C.1.8) (instead of (C.1.4)) gives

$$\begin{aligned} J_{r+1}(w) &= \phi(w)\Big[\psi^{(r)}(-w)\,\psi^{(r+1)}(w) + \psi^{(r)}(w)\,\psi^{(r+1)}(-w)\Big] = \\ &= \phi(w)\Big[\psi^{(r)}(-w)\big(w\,\psi^{(r)}(w) + r\,\psi^{(r-1)}(w)\big) + \\ &\qquad + \psi^{(r)}(w)\big(-w\,\psi^{(r)}(-w) + r\,\psi^{(r-1)}(-w)\big)\Big] = \\ &= r\,J_r(w) = \\ &= r!\,. \end{aligned} \tag{C.2.8}$$

This completes the proof. □

Bibliography

[1] R. Arratia, L. Goldstein, and L. Gordon, *Poisson approximation and the Chen-Stein method*, Statist. Sci., 5 (1990), pp. 403–434.

[2] A. D. Barbour, *Stein's method*, in Encyclopedia of Statistical Sciences, vol. 1, Wiley, New York, 1997, pp. 513–521.

[3] A. D. Barbour and L. H. Y. Chen, *An Introduction to Stein's Method*, vol. 4 of Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore, 2005.

[4] ——, *Stein's Method and Applications*, vol. 5 of Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore, 2005.

[5] A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*, vol. 2 of Oxford Studies in Probability, Oxford University Press, New York, 1992.

[6] A. D. Barbour, M. Karoński, and A. Ruciński, *A central limit theorem for decomposable random variables with applications to random graphs*, J. Combin. Theory Ser. B, 47 (1989), pp. 125–145.

[7] A. C. Berry, *The accuracy of the Gaussian approximation to the sum of independent variates*, Trans. Amer. Math. Soc., 49 (1941), pp. 122–136.

[8] E. Bolthausen, *An estimate of the remainder in a combinatorial central limit theorem*, Z. Wahrsch. verw. Gebiete, 66 (1984), pp. 379–386.

[9] L. H. Y. Chen, *Poisson approximation for dependent trials*, Ann. Probab., 3 (1975), pp. 534–545.

[10] L. H. Y. Chen, L. Goldstein, and A. Röllin, *Stein's method via induction*. ArXiv:1903.09319, 2019.

[11] L. H. Y. Chen and Q.-M. Shao, *A non-uniform Berry-Esseen bound via Stein's method*, Probab. Theory Related Fields, 120 (2001), pp. 236–254.

[12] ——, *Normal approximation under local dependence*, Ann. Probab., 32 (2004), pp. 1985–2028.

[13] D. L. Cohn, *Measure theory*, Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks], Birkhäuser/Springer, New York, second ed., 2013.

[14] P. R. de Montmort, *Essay de l'analyse sur les jeux de hazard*, Jacques Quillau, Paris, second ed., 1713.

[15] A. Dembo and Y. Rinott, *Some examples of normal approximations by Stein's method*, in Random discrete structures (Minneapolis, MN, 1993), vol. 76 of IMA Vol. Math. Appl., Springer, New York, 1996, pp. 25–44.

[16] C.-G. Esseen, *On the Liapounoff limit of error in the theory of probability*, Ark. Mat. Astr. Fys., 28A (1942), p. 19.

[17] C. G. Esseen, *A moment inequality with an application to the central limit theorem*, Skand. Aktuarietidskr., 39 (1956), pp. 160–170 (1957).

[18] S. N. Ethier and T. G. Kurtz, *Markov Processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986.

[19] L. Goldstein, *Berry–Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing*, J. Appl. Probab., 42 (2005), pp. 661–683.

[20] C. Heil, *Introduction to Real Analysis*, vol. 280 of Graduate Texts in Mathematics, Springer, Cham, 2019.

[21] P. L. Hsu, *The approximate distributions of the mean and variance of a sample of independent variables*, Ann. Math. Statistics, 16 (1945), pp. 1–29.

[22] V. Y. Korolev and I. G. Shevtsova, *An upper bound for the absolute constant in the Berry-Esseen inequality (Russian)*, Teor. Veroyatn. Primen., 54 (2009), pp. 671–695.

[23] L. Le Cam, *An approximation theorem for the Poisson binomial distribution*, Pacific J. Math., 10 (1960), pp. 1181–1197.

[24] M. Manetti, *Topology*, vol. 91 of Unitext, Springer, Cham, 2015. Translated from the 2014 Italian edition by Simon G. Chiossi, La Matematica per il 3+2.

[25] S. T. Rachev, *The Monge-Kantorovich problem on mass transfer and its applications in stochastics*, Theory Probab. Appl., 29 (1984), pp. 647–676.

[26] M. Raič, *CLT-related large deviation bounds based on Stein's method*, Adv. in Appl. Probab., 39 (2007), pp. 731–752.

[27] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes, and Martingales. Vol. 1*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000.

[28] I. G. Shevtsova, *On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands*. ArXiv:1111.6554, 2011.

[29] ——, *On the absolute constants in the Berry–Esseen inequality and its structural and nonuniform improvements (Russian)*, Inform. Primen., 7 (2013), pp. 124–125.

[30] T. B. Singh, *Introduction to Topology*, Springer, Singapore, 2019.

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

[32] ——, *Approximate Computation of Expectations*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7, Institute of Mathematical Statistics, Hayward, CA, 1986.

[33] T. Tao, *An Introduction to Measure Theory*, vol. 126 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2011.

[34] R. L. Wheeden and A. Zygmund, *Measure and Integral*, Pure and Applied Mathematics (Boca Raton), CRC Press, Boca Raton, FL, second ed., 2015. An introduction to real analysis.