Stein's density approach and information inequalities

Christophe Ley and Yvik Swan

April 4, 2013

Abstract

We provide a new perspective on Stein's so-called density approach by introducing a new operator and characterizing class which are valid for a much wider family of probability distributions on the real line. We prove an elementary factorization property of this operator and propose a new Stein identity which we use to derive information inequalities in terms of what we call the *generalized Fisher information distance*. We provide explicit bounds on the constants appearing in these inequalities for several important cases. We conclude with a comparison between our results and known results in the Gaussian case, hereby improving on several known inequalities from the literature.

1 Introduction

Charles Stein's crafty exploitation of the characterization

$$X \sim \mathcal{N}(0,1) \Longleftrightarrow \mathrm{E}\left[f'(X) - Xf(X)\right] = 0 \text{ for all bounded } f \in C^1(\mathbb{R}) \tag{1.1}$$

has given birth to a "method" which is now an acclaimed tool both in applied and in theoretical probability. The secret of the "method" lies in the structure of the operator $\mathcal{T}_\phi f(x) := f'(x) - xf(x)$ and in the flexibility in the choice of test functions f. For the origins we refer the reader to [39, 37, 36]; for an overview of the more recent achievements in this field we refer to the monographs [27, 3, 11] or the review articles [26, 30].

Among the many ramifications and extensions that the method has known, so far the connection with information theory has gone relatively unexplored. Indeed while it has long been known that Stein identities such as (1.1) are related to information theoretic tools and concepts (see, e.g., [19, 21, 13]), to the best of our knowledge the only references to explore this connection upfront are [4] in the context of compound Poisson approximation, and more recently [32, 31] for Poisson and Bernoulli approximation. In this paper and the companion paper [22] we extend Stein's characterization of the Gaussian (1.1) to a broad class of univariate distributions and, in doing so, provide an adequate framework in which the connection with information distances becomes transparent.

The structure of the present paper is as follows. In Section 2 we provide the new perspective on the density approach from [38] which allows to extend this construction to virtually any absolutely continuous probability distribution on the real line. In Section 3 we exploit the structure of our new operator to derive a family of Stein identities through which the connection with information distances becomes evident. In Section 4 we compute bounds on the constants appearing in our inequalities; our method of proof is, to the best of our knowledge, original. Finally in Section 5 we discuss specific examples.

2 The density approach

Let $\mathcal{G}$ be the collection of positive real functions $x \mapsto p(x)$ such that (i) their support $S_p := \{x \in \mathbb{R} : p(x) \text{ (exists and) is positive}\}$ is an interval with closure $\bar{S}_p = [a, b]$, for some $-\infty \le a < b \le \infty$, (ii) they are differentiable (in the usual sense) at every point in (a, b) with derivative $x \mapsto p'(x) := \frac{d}{dy}p(y)|_{y=x}$ and (iii) $\int_{S_p} p(y)dy = 1$. Obviously, each $p \in \mathcal{G}$ is the density (with respect to the Lebesgue measure) of an absolutely continuous random variable. Throughout we adopt the convention

$$\frac{1}{p(x)} = \begin{cases} \frac{1}{p(x)} & \text{if } x \in S_p \\ 0 & \text{otherwise;} \end{cases}$$

this implies, in particular, that $p(x)/p(x) = \mathbb{I}_{S_p}(x)$, the indicator function of the support S_p. As final notation, for $p \in \mathcal{G}$ we write $\mathrm{E}_p[l(X)] := \int_{S_p} l(x)p(x)dx$.

With this setup in hand we are ready to provide the two main definitions of this paper (namely, a class of functions and an operator) and to state and prove our first main result (namely, a characterization).

Definition 2.1. *To $p \in \mathcal{G}$ we associate (i) the collection $\mathcal{F}(p)$ of functions $f : \mathbb{R} \to \mathbb{R}$ such that the mapping $x \mapsto f(x)p(x)$ is differentiable on the interior of S_p and $f(a^+)p(a^+) = f(b^-)p(b^-) = 0$, and (ii) the operator*

$\mathcal{T}_p : \mathcal{F}(p) \to \mathbb{R}^\star : f \mapsto \mathcal{T}_p f$ *defined through*

$$\mathcal{T}_p f : \mathbb{R} \to \mathbb{R} : x \mapsto \mathcal{T}_p f(x) := \frac{1}{p(x)} \frac{d}{dy}(f(y)p(y))\Big|_{y=x}. \tag{2.1}$$

We call $\mathcal{F}(p)$ *the class of* test functions *associated with* p*, and* $\mathcal{T}_p$ *the* Stein operator *associated with* p*.*

Theorem 2.1. *Let* $p, q \in \mathcal{G}$ *and let* $Q(b) = \int_a^b q(u)du$*. Then* $\int_{-\infty}^{+\infty} \mathcal{T}_p f(y) q(y) dy = 0$ *for all* $f \in \mathcal{F}(p)$ *if, and only if,* $q(x) = p(x)Q(b)$ *for all* $x \in S_p$*.*

Proof. If $Q(b) = 0$ the statement holds trivially. We now take $Q(b) > 0$. To see the sufficiency, note that the hypotheses on f, p and q guarantee that

$$\begin{aligned}\int_{-\infty}^{\infty} \mathcal{T}_p f(y) q(y) dy &= Q(b) \int_a^b \frac{d}{du}(f(u)p(u))|_{u=y} dy \\ &= Q(b) \left(f(b^-)p(b^-) - f(a^+)p(a^+) \right) = 0.\end{aligned}$$

To see the necessity, first note that the condition $\int_{\mathbb{R}} \mathcal{T}_p f(y) q(y) dy = 0$ implies that the function $y \mapsto \mathcal{T}_p f(y) q(y)$ be Lebesgue-integrable. Next define for $z \in \mathbb{R}$ the function

$$l_z(u) := (\mathbb{I}_{(a,z]}(u) - P(z))\mathbb{I}_{S_p}(u)$$

with $P(z) := \int_a^z p(u)du$, which satisfies

$$\int_a^b l_z(u)p(u)du = 0.$$

Then the function

$$f_z^p(x) := \frac{1}{p(x)} \int_a^x l_z(u)p(u)du \left(= -\frac{1}{p(x)} \int_x^b l_z(u)p(u)du \right)$$

belongs to $\mathcal{F}(p)$ for all z and satisfies the equation

$$\mathcal{T}_p f_z^p(x) = l_z(x)$$

for all $x \in S_p$. For this choice of test function we then obtain

$$\int_{-\infty}^{+\infty} \mathcal{T}_p f_z^p(y) q(y) dy = \int_{-\infty}^{+\infty} l_z(y) q(y) dy = (Q(z) - P(z)Q(b))\mathbb{I}_{S_p}(z),$$

with $Q(z) := \int_a^z q(u)du$. Since this integral equals zero by hypothesis, it follows that $Q(z) = P(z)Q(b)$ for all $z \in S_p$, hence the claim holds. □

The above is, in a sense, nothing more than a peculiar statement of what is often referred to as a "Stein characterization". Within the more conventional framework of real random variables having absolutely continuous densities, Theorem 2.1 reads as follows.

Corollary 2.1 (The density approach)**.** *Let X be an absolutely continuous random variable with density $p \in \mathcal{G}$. Let Y be another absolutely continuous random variable. Then $\mathrm{E}\left[\mathcal{T}_p f(Y)\right] = 0$ for all $f \in \mathcal{F}(p)$ if, and only if, either $\mathrm{P}(Y \in S_p) = 0$ or $\mathrm{P}(Y \in S_p) > 0$ and*

$$\mathrm{P}\left(Y \leq z \,|\, Y \in S_p\right) = \mathrm{P}(X \leq z)$$

for all $z \in S_p$.

Corollary 2.1 extends the density approach from [38] or [10, 11] to a much wider class of distributions; it also contains the Stein characterizations for the Pearson given in [33] and the more recent general characterizations studied in [14, 17]. There is, however, a significant shift operated between our "derivative of a product" operator (2.1) and the standard way of writing these operators in the literature. Indeed, while one can always distribute the derivative in (2.1) to obtain (at least formally) the expansion

$$\mathcal{T}_p f(x) = \left(f'(x) + \frac{p'(x)}{p(x)} f(x) \right) \mathbb{I}_{S_p}(x), \tag{2.2}$$

the latter requires f be differentiable on S_p in order to make sense. We do not require this, neither do we require that each summand in (2.2) be well-defined on S_p nor do we need to impose integrability conditions on f for Theorem 2.1 (and thus Corollary 2.1) to hold! Rather, our definition of $\mathcal{F}(p)$ allows to identify a collection of minimal conditions on the class of test functions f for the resulting operator $\mathcal{T}_p$ to be orthogonal to p w.r.t. the Lebesgue measure, and thus characterize p.

Example 2.1. *Take $p = \phi$, the standard Gaussian. Then $\mathcal{F}(\phi)$ is composed of all real-valued functions f such that (i) $x \mapsto f(x)e^{-x^2/2}$ is differentiable on $\mathbb{R}$ and (ii) $\lim_{x \to \pm\infty} f(x)e^{-x^2/2} = 0$. In particular $\mathcal{F}(\phi)$ contains the collection of all differentiable bounded functions and*

$$\mathcal{T}_\phi f(x) = f'(x) - x f(x),$$

which is Stein's well-known operator for characterizing the Gaussian (see, e.g., [36, 3, 11]). There are of course many other subclasses that can be of *interest. For example the class $\mathcal{F}(\phi)$ also contains the collection of functions $f(x) = -f_0'(x)$ with f_0 a twice differentiable bounded function; for these we get*

$$\mathcal{T}_\phi f(x) = x f_0'(x) - f_0''(x),$$

the generator of an Ornstein-Uhlenbeck process, see [2, 18, 27]. The class $\mathcal{F}(\phi)$ as well contains the collection of functions of the form $f(x) = H_n(x) f_0(x)$ for H_n the n-th Hermite polynomial and f_0 any differentiable and bounded function. For these f we get

$$\mathcal{T}_\phi f(x) = H_n(x) f_0'(x) - H_{n+1}(x) f_0(x),$$

an operator already discussed in [16] (equation (38)).

Example 2.2. *Take $p = Exp$ the standard rate-one exponential distribution. Then $\mathcal{F}(Exp)$ is composed of all real-valued functions f such that (i) $x \mapsto f(x)e^{-x}$ is differentiable on $(0, +\infty)$, (ii) $f(0) = 0$ and (iii) $\lim_{x\to+\infty} f(x)e^{-x} = 0$. In particular $\mathcal{F}(Exp)$ contains the collection of all differentiable bounded functions such that $f(0) = 0$ and*

$$\mathcal{T}_{Exp} f(x) = \left(f'(x) - f(x)\right) \mathbb{I}_{[0,\infty)}(x),$$

the operator usually associated to the exponential, see [24, 28, 38]. The class $\mathcal{F}(Exp)$ also contains the collection of functions of the form $f(x) = x f_0(x)$ for f_0 any differentiable bounded function. For these f we get

$$\mathcal{T}_{Exp} f(x) = \left(x f_0'(x) + (1-x) f_0(x)\right) \mathbb{I}_{[0,\infty)}(x),$$

an operator put to use in [9].

Example 2.3. *Finally take $p = Beta(\alpha, \beta)$ the beta distribution with parameters $(\alpha, \beta) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+$. Then $\mathcal{F}(Beta(\alpha, \beta))$ is composed of all real-valued functions f such that (i) $x \mapsto f(x)x^{\alpha-1}(1-x)^{\beta-1}$ is differentiable on $(0,1)$, (ii) $\lim_{x\to 0} f(x)x^{\alpha-1}(1-x)^{\beta-1} = 0$ and (iii) $\lim_{x\to 1} f(x)x^{\alpha-1}(1-x)^{\beta-1} = 0$. In particular $\mathcal{F}(Beta(\alpha, \beta))$ contains the collection of functions of the form $f(x) = (x(1-x)) f_0(x)$ with f_0 any differentiable bounded function. For these f we get*

$$\mathcal{T}_{Beta(\alpha,\beta)} f(x) = \left(\left(\alpha(1-x) - \beta x\right) f_0(x) + x(1-x) f_0'(x)\right) \mathbb{I}_{[0,1]}(x),$$

an operator recently put to use in, e.g., [17, 14].

There are obviously many more distributions that can be tackled as in the previous examples (including the Pearson case from [33]), which we leave to the interested reader.

3 Stein-type identities and the generalized Fisher information distance

It has long been known that, in certain favorable circumstances, the properties of the Fisher information or of the Shannon entropy can be used quite effectively to prove information theoretic central limit theorems; the early references in this vein are [35, 6, 5, 23]. Convergence in information CLTs is generally studied in terms of information (pseudo-)distances such as the *Kullback-Leibler divergence* between two densities p and q, defined as

$$d_{\mathrm{KL}}(p||q) = \mathrm{E}_q\left[\log\left(\frac{q(X)}{p(X)}\right)\right], \tag{3.1}$$

or the *Fisher information distance*

$$\mathcal{J}(\phi, q) = \mathrm{E}_q\left[\left(X + \frac{q'(X)}{q(X)}\right)^2\right] \tag{3.2}$$

which measures deviation between any density q and the standard Gaussian ϕ. Though they allow for extremely elegant proofs, convergence in the sense of (3.1) or (3.2) results in very strong statements. Indeed both (3.1) and (3.2) are known to dominate more "traditional" probability metrics. More precisely we have, on the one hand, *Pinsker's inequality*

$$d_{\mathrm{TV}}(p, q) \le \frac{1}{\sqrt{2}}\sqrt{d_{\mathrm{KL}}(p||q)}, \tag{3.3}$$

for $d_{\mathrm{TV}}(p, q)$ the *total variation distance* between the laws p and q (see, e.g., [15, p. 429]), and, on the other hand,

$$d_{L^1}(\phi, q) \le \sqrt{2}\sqrt{\mathcal{J}(\phi, q)} \tag{3.4}$$

for $d_{L^1}(\phi, q)$ the L^1 *distance* between the laws ϕ and q (see [20, Lemma 1.6]). These information inequalities show that convergence in the sense of (3.1) or (3.2) implies convergence in total variation or in L^1, for example. Note that one can further use De Brujn's identity on (3.3) to deduce that convergence in Fisher information is itself stronger than convergence in relative entropy.

While Pinsker's inequality (3.3) is valid irrespective of the choice of p and q (and enjoys an extension to discrete random variables), both (3.2) and (3.4) are reserved for Gaussian convergence. Now there exist extensions of the distance (3.2) to non-Gaussian distributions (see [4] for the discrete case) which, as could be expected, have also been shown to dominate the more traditional probability metrics. There is, however, no general counterpart of Pinsker's inequality for the Fisher information distance (3.2); at least there exists, to the best of our knowledge, no inequality in the literature which extends (3.4) to a general couple of densities p and q.

In this section we use the density approach outlined in Section 2 to construct Stein-type identities which provide the required extension of (3.4). More precisely, we will show that a wide family of probability metrics (including the *Kolmogorov*, the *Wasserstein* and the L^1 distances) is dominated by the quantity

$$\mathcal{J}(p,q) := \mathrm{E}_q\left[\left(\frac{p'(X)}{p(X)} - \frac{q'(X)}{q(X)}\right)^2\right]. \tag{3.5}$$

Our bounds, moreover, contain an explicit constant which will be shown in Section 4 to be at worst as good as the best bounds in all known instances. In the spirit of [4] we call (3.5) the *generalized Fisher information distance* between the densities p and q, although here we slightly abuse of language since (3.5) rather defines a pseudo-distance than a *bona fide* metric between probability density functions.

We start with an elementary statement which relates, for $p \neq q$, the Stein operators $\mathcal{T}_p$ and $\mathcal{T}_q$ through the difference of their respective *score functions* $\frac{p'}{p}$ and $\frac{q'}{q}$.

Lemma 3.1. *Let p and q be probability density functions in $\mathcal{G}$ with respective supports S_p and S_q. Let $S_q \subseteq S_p$ and define*

$$r(p,q)(x) := \left(\frac{p'(x)}{p(x)} - \frac{q'(x)}{q(x)}\right)\mathbb{I}_{S_p}(x).$$

Suppose that $\mathcal{F}(p) \cap \mathcal{F}(q) \neq \emptyset$. Then, for all $f \in \mathcal{F}(p) \cap \mathcal{F}(q)$, we have

$$\mathcal{T}_p f(x) = \mathcal{T}_q f(x) + f(x) r(p,q)(x) + \mathcal{T}_p f(x)\mathbb{I}_{S_p \setminus S_q}(x),$$

and therefore

$$\mathrm{E}_q\left[\mathcal{T}_p f(X)\right] = \mathrm{E}_q\left[f(X) r(p,q)(X)\right]. \tag{3.6}$$

Proof. Splitting S_p into $S_q \cup \{S_p \setminus S_q\}$, we have

$$f(y)p(y) = f(y)q(y)p(y)/q(y)\mathbb{I}_{S_q}(y) + f(y)p(y)\mathbb{I}_{S_p\setminus S_q}(y)$$

for any real-valued function f. At any x in the interior of S_p we thus can write

$$\begin{aligned}
&\mathcal{T}_p f(x) \\
&= \frac{\frac{d}{dy}(f(y)q(y)p(y)/q(y))\Big|_{y=x}}{p(x)} \mathbb{I}_{S_q}(x) + \mathcal{T}_p f(x) \mathbb{I}_{S_p \setminus S_q}(x) \\
&= \frac{\frac{d}{dy}(f(y)q(y))\Big|_{y=x}}{p(x)} \frac{p(x)}{q(x)} + f(x)q(x) \frac{\frac{d}{dy}(p(y)/q(y))\Big|_{y=x}}{p(x)} + \mathcal{T}_p f(x) \mathbb{I}_{S_p \setminus S_q}(x) \\
&= \mathcal{T}_q f(x) + f(x) \frac{q(x)}{p(x)} \frac{d}{dy}(p(y)/q(y))\Big|_{y=x} + \mathcal{T}_p f(x) \mathbb{I}_{S_p \setminus S_q}(x).
\end{aligned}$$

The first claim readily follows by simplification, the second by taking expectations under q which cancels the first term $\mathcal{T}_q f(x)$ (by definition) as well as the third term $\mathcal{T}_p f(x) \mathbb{I}_{S_p \setminus S_q}(x)$ (since the supports do not coincide). □

Remark 3.1. *Our proof of Lemma 3.1 may seem circumvoluted; indeed a much easier proof is obtainable by writing $\mathcal{T}_p$ under the form (2.2). We nevertheless stick to the "derivative of a product" structure of our operator because this dispenses us with superfluous – and, in some cases, unwanted – differentiability conditions on the test functions.*

From identity (3.6) we deduce the following immediate result, which requires no proof.

Lemma 3.2. *Let p and q be probability density functions in $\mathcal{G}$ with respective supports $S_q \subseteq S_p$. Let l be a real-valued function such that $\mathrm{E}_p[l(X)]$ and $\mathrm{E}_q[l(X)]$ exist; also suppose that there exists $f \in \mathcal{F}(p) \cap \mathcal{F}(q)$ such that*

$$\mathcal{T}_p f(x) = (l(x) - \mathrm{E}_p[l(X)]) \mathbb{I}_{S_p}(x); \tag{3.7}$$

we denote this function f_l^p. Then

$$\mathrm{E}_q[l(X)] - \mathrm{E}_p[l(X)] = \mathrm{E}_q[f_l^p(X) r(p,q)(X)]. \tag{3.8}$$

The identity (3.8) belongs to the family of so-called "Stein-type identities" discussed for instance in [16, 7, 1]. In order to be of use, such identities need to be valid over a large class of test functions l. Now it is immediate to write out the solution f_l^p of the so-called "Stein equation" (3.7) explicitly for any given p and l; it is therefore relatively simple to identify under which conditions on l and q the requirement $f_l^p \in \mathcal{F}(q)$ is verified (since $f_l^p \in \mathcal{F}(p)$ is anyway true).

Remark 3.2. *For instance, for $p = \phi$ the standard Gaussian, one easily sees that $\lim_{x\to\pm\infty} f_l^{\phi}(x) = 0$, hence, when $S_q = S_\phi = \mathbb{R}$, q only has to be (differentiable and) bounded for f_l^{ϕ} to belong to $\mathcal{F}(q)$. However, when $S_q \subset \mathbb{R}$, then q has to satisfy, moreover, the stronger condition of vanishing at the endpoints of its support S_q since f_l^{ϕ} needs not equal zero on any finite points in $\mathbb{R}$.*

We shall see in the next section that the required conditions for $f_l^p \in \mathcal{F}(q)$ are satisfied in many important cases by wide classes of functions l. The resulting flexibility makes (3.8) a surprisingly powerful identity, as can be seen from our next result.

Theorem 3.1. *Let p and q be probability density functions in $\mathcal{G}$ with respective supports $S_q \subseteq S_p$ and such that $\mathcal{F}(p) \cap \mathcal{F}(q) \neq \emptyset$. Let*

$$d_{\mathcal{H}}(p,q) = \sup_{l\in\mathcal{H}} |\mathrm{E}_q[l(X)] - \mathrm{E}_p[l(X)]| \tag{3.9}$$

for some class of functions $\mathcal{H}$. Suppose that for all $l \in \mathcal{H}$ the function f_l^p, as defined in (3.7), exists and satisfies $f_l^p \in \mathcal{F}(p) \cap \mathcal{F}(q)$. Then

$$d_{\mathcal{H}}(p,q) \leq \kappa_{\mathcal{H}}^p \sqrt{\mathcal{J}(p,q)}, \tag{3.10}$$

where

$$\kappa_{\mathcal{H}}^p = \sup_{l\in\mathcal{H}} \sqrt{\mathrm{E}_q[(f_l^p(X))^2]} \tag{3.11}$$

and

$$\mathcal{J}(p,q) = \mathrm{E}_q[(r(p,q)(X))^2], \tag{3.12}$$

the generalized Fisher information distance between the densities p and q.

This theorem implies that all probability metrics that can be written in the form (3.9) are bounded by the generalized Fisher information distance $\mathcal{J}(p,q)$ (which, of course, can be infinite for certain choices of p and q). Equation (3.10) thus represents the announced extension of (3.4) to any couple of densities (p,q) and hence constitutes, in a sense, a counterpart to Pinsker's inequality (3.3) for the Fisher information distance. We will see in Section 5 how this inequality reads for specific choices of $\mathcal{H}$, p and q.

4 Bounding the constants

The constants $\kappa_{\mathcal{H}}^p$ in (3.11) depend on both densities p and q and therefore, to be fair, should be denoted $\kappa_{\mathcal{H}}^{p,q}$. Our notation is nevertheless justified because we always have

$$\kappa_{\mathcal{H}}^p \le \sup_{l\in\mathcal{H}} \|f_l^p\|_\infty, \tag{4.1}$$

where the latter bounds (sometimes referred to as *Stein factors* or *magic factors*) do not depend on q and have been computed for many choices of $\mathcal{H}$ and p. Consequently, $\kappa_{\mathcal{H}}^p$ is finite in many known cases – including, of course, that of a Gaussian target.

Example 4.1. *Take $p = \phi$, the standard Gaussian. Then, from (4.1), we get the bounds (i) $\kappa_{\mathcal{H}}^p \le \sqrt{\pi/2}$ for $\mathcal{H}$ the collection of Borel functions in $[0,1]$ (see [27, Theorem 3.3.1]); (ii) $\kappa_{\mathcal{H}}^p \le \sqrt{2\pi}/4$ for $\mathcal{H}$ the class of indicator functions for lower half-lines (see [27, Theorem 3.4.2]); and (iii) $\kappa_{\mathcal{H}}^p \le \sqrt{\pi/2}\sup_{l\in\mathcal{H}}\min\left(\|l - \mathrm{E}_p\left[l(X)\right]\|_\infty, 2\|l'\|_\infty\right)$ for $\mathcal{H}$ the class of absolutely continuous functions on $\mathbb{R}$ (see [12, Lemma 2.3]). See also [3, 11, 27, 29] for more examples.*

Bounds such as (4.1) are sometimes too rough to be satisfactory. We now provide an alternative bound for $\kappa_{\mathcal{H}}^p$ which, remarkably, improves upon the best known bounds even in well-trodden cases such as the Gaussian. We focus on target densities of the form

$$p(x) = ce^{-d|x|^\alpha}\mathbb{I}_S(x), \quad \alpha \ge 1, \tag{4.2}$$

with S a scale-invariant subset of $\mathbb{R}$ (that is, either $\mathbb{R}$ or the open/closed positive/negative real half lines), $d > 0$ some constant and c the appropriate normalizing constant. The exponential, the Gaussian or the limit distribution for the Ising model on the complete graph from [10] are all of the form (4.2). Of course, for $S = \mathbb{R}$, (4.2) represents power exponential densities.

Theorem 4.1. *Take $p \in \mathcal{G}$ as in (4.2) and $q \in \mathcal{G}$ such that $S_q = S$. Consider $h : \mathbb{R} \to \mathbb{R}$ some Borel function with p-mean $\mathrm{E}_p\left[h(X)\right] = 0$. Let f_h^p be the unique bounded solution of the Stein equation*

$$\mathcal{T}_p f(x) = h(x). \tag{4.3}$$

Then

$$\sqrt{\mathrm{E}_q\left[\left(f_h^p(X)\right)^2\right]} \le \frac{||h||_\infty}{2^{\frac{1}{\alpha}}}. \tag{4.4}$$

Proof. Under the assumption that $\mathrm{E}_p[h(X)] = 0$, the unique bounded solution of (4.3) is given by

$$f_h^p(x) = \begin{cases} \dfrac{1}{p(x)}\displaystyle\int_{-\infty}^x h(y)p(y)dy & \text{if } x \le 0, \\ \dfrac{-1}{p(x)}\displaystyle\int_x^\infty h(y)p(y)dy & \text{if } x \ge 0, \end{cases}$$

the function being, of course, put to 0 if x is outside the support of p. Then

$$\begin{aligned}\mathrm{E}_q\left[(f_h^p(X))^2\right] &= \int_{-\infty}^0 q(x)\left(\frac{1}{p(x)}\int_{-\infty}^x h(y)p(y)dy\right)^2 dx\\ &\quad+\int_0^\infty q(x)\left(\frac{1}{p(x)}\int_x^\infty h(y)p(y)dy\right)^2 dx\\ &=: I^- + I^+,\end{aligned}$$

where $I^- = 0$ (resp., $I^+ = 0$) if $\bar{S} = \mathbb{R}^+$ (resp., $\bar{S} = \mathbb{R}^-$).

We first tackle I^-. Setting $p(x) = ce^{-d|x|^\alpha}\mathbb{I}_S(x)$ and using Jensen's inequality, we get

$$\begin{aligned}I^- &= \int_{-\infty}^0 q(x)\left(e^{d|x|^\alpha}\int_{-\infty}^x h(u)e^{-d|u|^\alpha}du\right)^2 dx\\ &\leq \int_{-\infty}^0 q(x)\left(e^{d|x|^\alpha}\int_{-\infty}^x |h(u)|e^{-d|u|^\alpha}du\right)^2 dx\\ &\leq \int_{-\infty}^0 q(x)\left(e^{2d|x|^\alpha}\int_{-\infty}^x h^2(u)e^{-2d|u|^\alpha}du\right) dx\\ &= \frac{1}{2^{1/\alpha}}\int_{-\infty}^0 q(x)\left(e^{2d|x|^\alpha}\int_{-\infty}^{2^{1/\alpha}x} h^2(u/2^{1/\alpha})e^{-d|u|^\alpha}du\right) dx,\end{aligned}$$

where the last equality follows from a simple change of variables. Applying Hölder's inequality we obtain

$$I^- \leq \frac{\gamma_q^{1/2}}{2^{1/\alpha}}\sqrt{\int_{-\infty}^0 q(x)\left(e^{2d|x|^\alpha}\int_{-\infty}^{2^{1/\alpha}x} h^2(u/2^{1/\alpha})e^{-d|u|^\alpha}du\right)^2 dx} =: I_1^-,$$

where $\gamma_q = \mathrm{P}_q(X < 0) := \int_{-\infty}^0 q(x)dx$. Repeating the *Jensen's inequality-change of variables-Hölder's inequality* scheme once more yields

$$I^- \leq I_1^- \leq I_2^-$$

with

$$I_2^- = \frac{\gamma_q^{\frac{1}{2}+\frac{1}{4}}}{2^{\frac{1}{\alpha}(1+\frac{1}{2})}}\left(\int_{-\infty}^0 q(x)\left(e^{4d|x|^\alpha}\int_{-\infty}^{(2^{1/\alpha})^2x} h^4\left(\frac{u}{(2^{1/\alpha})^2}\right)e^{-d|u|^\alpha}du\right)^2 dx\right)^{\frac{1}{4}}.$$

Iterating this procedure $m \in \mathbb{N}$ times we deduce

$$I^- \leq I_1^- \leq \ldots \leq I_m^-$$

with I_m^- given by

$$\frac{\gamma_q^{N(m)-1}}{2^{\frac{1}{\alpha}N(m)}}\left(\int_{-\infty}^{0} q(x)\left(e^{2^m d|x|^\alpha}\int_{-\infty}^{(2^{1/\alpha})^m x} h^{2^m}\left(\frac{u}{(2^{1/\alpha})^m}\right)e^{-d|u|^\alpha}du\right)^2 dx\right)^{\frac{1}{2^m}},$$

where $N(m) = 1 + \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^m}$. Bounding $h^{2^m}\left(\frac{u}{(2^{1/\alpha})^m}\right)$ by $(||h||_\infty)^{2^m}$ simplifies the above into

$$\frac{(||h||_\infty)^2\gamma_q^{N(m)-1}}{2^{\frac{1}{\alpha}N(m)}}\left(\int_{-\infty}^{0} q(x)\left(e^{2^m d|x|^\alpha}\int_{-\infty}^{(2^{1/\alpha})^m x} e^{-d|u|^\alpha}du\right)^2 dx\right)^{\frac{1}{2^m}}.$$

Since the mapping $y \mapsto \eta(y) := e^{d|y|^\alpha}\int_{-\infty}^{y} e^{-d|u|^\alpha}du$ attains its maximal value at 0 for $\alpha \geq 1$ (indeed,

$$\begin{aligned}\eta'(y) &= 1 - e^{d|y|^\alpha} d\,\alpha|y|^{\alpha-1}\int_{-\infty}^{y} e^{-d|u|^\alpha}du \\ &\geq 1 - e^{d|y|^\alpha}\int_{-\infty}^{y} d\alpha|u|^{\alpha-1}e^{-d|u|^\alpha}du = 0,\end{aligned}$$

hence η is monotone increasing), the interior of the parenthesis becomes

$$\int_{-\infty}^{0} q(x)\left(e^{2^m d|x|^\alpha}\int_{-\infty}^{(2^{1/\alpha})^m x} e^{-d|u|^\alpha}du\right)^2 dx \leq \int_{-\infty}^{0} q(x)\frac{1}{c^2}dx = \frac{\gamma_q}{c^2}.$$

Note that here we have used, for any support S, $\int_{-\infty}^{0} ce^{-d|u|^\alpha}du \leq 1$. Elevated to the power $1/(2m)$, this factor tends to 1 as $m \to \infty$. Since we also have $\lim_{m\to\infty} N(m) = 2$ we finally obtain

$$I^- \leq \lim_{m\to\infty} I_m^- \leq \frac{(||h||_\infty)^2}{2^{\frac{2}{\alpha}}}\mathrm{P}_q(X < 0).$$

Similar manipulations allow to bound I^+ by $\frac{(||h||_\infty)^2}{2^{\frac{2}{\alpha}}}\mathrm{P}_q(X > 0)$. Combining both bounds then allows us to conclude that

$$\sqrt{\mathrm{E}_q\left[(f_h^p(X))^2\right]} \leq \frac{||h||_\infty}{2^{\frac{1}{\alpha}}},$$

hence the claim holds. $\square$

This result of course holds true without worrying about $f_h^p \in \mathcal{F}(q)$. However, in order to make use of these bounds in the present context, the latter condition has to be taken care of. For densities of the form (4.2), one easily sees that $f_h^p \in \mathcal{F}(q)$ for all (differentiable and) bounded densities q for $\alpha > 1$, with the additional assumption, for $\alpha = 1$, that $\lim_{x\to\pm\infty} q(x) = 0$.

Example 4.2. *Take $p = \phi$, the standard Gaussian. Then, from (4.4),*

$$\kappa_{\mathcal{H}}^p \leq \frac{1}{\sqrt{2}} \sup_{l\in\mathcal{H}} \|l - \mathrm{E}_\phi\left[l(X)\right]\|_\infty. \tag{4.5}$$

Comparing with the bounds from Example 4.1 we see that (4.5) significantly improves on the constants in cases (i) and (iii); it is slightly worse in case (ii).

5 Applications

A wide variety of probability distances can be written under the form (3.9). For instance the total variation distance is given by

$$d_{\mathrm{TV}}(p, q) = \sup_{A\subset\mathbb{R}} \left| \int_A (p(x) - q(x))dx \right| = \frac{1}{2} \sup_{h\in\mathcal{H}_{B[-1,1]}} \left|\mathrm{E}_p\left[h(X)\right] - \mathrm{E}_q\left[h(X)\right]\right|$$

with $\mathcal{H}_{B[-1,1]}$ the class of Borel functions in $[-1, 1]$, the Wasserstein distance is given by

$$d_{\mathrm{W}}(p, q) = \sup_{h\in\mathcal{H}_{\mathrm{Lip1}}} \left|\mathrm{E}_p\left[h(X)\right] - \mathrm{E}_q\left[h(X)\right]\right|$$

with $\mathcal{H}_{\mathrm{Lip1}}$ the class of Lipschitz-1 functions on $\mathbb{R}$ and the Kolmogorov distance is given by

$$d_{\mathrm{Kol}}(p, q) = \sup_{z\in\mathbb{R}} \left| \int_{-\infty}^z (p(x) - q(x))dx \right| = \sup_{h\in\mathcal{H}_{HL}} \left|\mathrm{E}_p\left[h(X)\right] - \mathrm{E}_q\left[h(X)\right]\right|$$

with $\mathcal{H}_{HL}$ the class of indicators of lower half lines. We refer to [15] for more examples and for an interesting overview of the relationships between these probability metrics.

Specifying the class $\mathcal{H}$ in Theorem 3.1 allows to bound all such probability metrics in terms of the generalized Fisher information distance (3.12). It remains to compute the constant (3.11), which can be done for all p of the form (4.2) through (4.4). The following result illustrates these computations in several important cases.

Corollary 5.1. *Take $p \in \mathcal{G}$ as in (4.2) and $q \in \mathcal{G}$ such that $S_q = S$. For $\alpha > 1$, suppose that q is (differentiable and) bounded over S; for $\alpha = 1$, assume moreover that q vanishes at the infinite endpoint(s) of S. Then we have the following inequalities:*

1. $$d_{\mathrm{TV}}(p, q) \leq 2^{-\frac{1}{\alpha}} \sqrt{\mathcal{J}(p, q)}$$

2. $$d_{\mathrm{Kol}}(p, q) \leq 2^{-\frac{1}{\alpha}} \sqrt{\mathcal{J}(p, q)}$$

3. $$d_{\mathrm{W}}(p, q) \leq \frac{\sup_{l \in \mathcal{H}_{\mathrm{Lip1}}} ||l - \mathrm{E}_p[l(X)]||_\infty}{2^{\frac{1}{\alpha}}} \sqrt{\mathcal{J}(p, q)}$$

4. $$d_{L^1}(p, q) = \int_S |p(x) - q(x)| dx \leq 2^{1-\frac{1}{\alpha}} \sqrt{\mathcal{J}(p, q)}.$$

If, for all $y \in S$, q is such that the function $f_l^p(x) = e^{d|x|^\alpha}(\mathbb{I}_{[y,b)}(x) - P(x))$, where P denotes the cumulative distribution function associated with p, belongs to $\mathcal{F}(q)$, then

$$d_{\sup}(p, q) = \sup_{x \in \mathbb{R}} |p(x) - q(x)| \leq \sqrt{\mathcal{J}(p, q)}.$$

Proof. The first three points follow immediately from the definition of the distances and Theorems 3.1 and 4.1. To show the fourth, note that

$$\int_S |p(x) - q(x)| dx = \mathrm{E}_p[l(X)] - \mathrm{E}_q[l(X)]$$

for $l(u) = \mathbb{I}_{[p(u) \geq q(u)]} - \mathbb{I}_{[q(u) \geq p(u)]} = 2\mathbb{I}_{[p(u) \geq q(u)]} - 1$. For the last case note that

$$d_{\sup}(p, q) := \sup_{y \in S} |p(y) - q(y)| = \sup_{y \in S} |\mathrm{E}_p[l_y(X) - \mathrm{E}_q[l_y(X)]|$$

for $l_y(x) = \delta_{\{x=y\}}$ the Dirac delta function in $y \in S$. The computation of the constant $\kappa_{\mathcal{H}}^p$ in this case requires a different approach from our Theorem 4.1. We defer this to the Appendix. □

We conclude this section, and the paper, with explicit computations in the Gaussian case $p = \phi$, hence for the classical Fisher information distance. From here on we adopt the more standard notations and write $\mathcal{J}(X)$ instead of $\mathcal{J}(\phi, q)$, for X a random variable with density q (which has support $\mathbb{R}$). Immediate applications of the above yield

$$\int_S |\phi(x) - q(x)|\, dx \le \sqrt{2}\sqrt{\mathcal{J}(X)},$$

which is the second inequality in [20, Lemma 1.6] (obtained by entirely different means). Similarly we readily deduce

$$\sup_{x \in \mathbb{R}} |\phi(x) - q(x)| \le \sqrt{\mathcal{J}(X)};$$

this is a significant improvement on the constant in [20, 35].

Next further suppose that X has density q with mean μ and variance σ^2. Take $Z \sim p$ with $p = \phi_{\mu_0, \sigma_0^2}$, the Gaussian with mean μ_0 and variance σ_0^2. Then

$$\mathcal{J}(X) = \mathrm{E}_q \left[\left(\frac{q'(X)}{q(X)} + \frac{X - \mu_0}{\sigma_0^2} \right)^2 \right] = I(X) + \frac{(\mu - \mu_0)^2}{\sigma_0^4} + \frac{1}{\sigma_0^2} \left(\frac{\sigma^2}{\sigma_0^2} - 2 \right),$$

where $I(X) = \mathrm{E}_q \left[(q'(X)/q(X))^2 \right]$ is the Fisher information of the random variable X. General bounds are thus also obtainable from (3.10) in terms of

$$\Psi := \Psi(\mu, \mu_0, \sigma, \sigma_0) = \frac{(\mu - \mu_0)^2}{\sigma_0^4} + \frac{1}{\sigma_0^2} \left(\frac{\sigma^2}{\sigma_0^2} - 1 \right).$$

and the quantity

$$\Gamma(X) = I(X) - \frac{1}{\sigma_0^2},$$

referred to as the *Cramér-Rao functional* for q in [25]. In particular, we deduce from Theorem 4.1 and the definition of the total variation distance that

$$d_{\mathrm{TV}}(\phi_{\mu_0, \sigma_0^2}, q) \le \frac{1}{\sqrt{2}} \sqrt{\Gamma(X) + \Psi}.$$

This is an improvement (in the constant) on [25, Lemma 3.1], and is also related to [8, Corollary 1.1]. Similarly, taking $\mathcal{H}$ the collection of indicators for lower half lines we can use (4.1) and the bounds from [12, Lemma 2.2] to deduce

$$d_{\mathrm{Kol}}(\phi_{\mu_0, \sigma_0^2}, q) \le \frac{\sqrt{2\pi}}{4} \sigma_0 \sqrt{\Gamma(X) + \Psi}.$$

Further specifying $q = \phi_{\mu_1,\sigma_1^2}$ we see that

$$\sigma_0\sqrt{\Gamma(X)+\Psi} \le \frac{\left|\sigma_1^2-\sigma_0^2\right|}{\sigma_0\sigma_1} + \frac{|\mu_1-\mu_0|}{\sigma_0},$$

to be compared with [27, Proposition 3.6.1]. Lastly take $Z \sim \phi$ the standard Gaussian and $X \stackrel{d}{=} F(Z)$ for F some monotone increasing function on $\mathbb{R}$ such that $f = F'$ is defined everywhere. Then straightforward computations yield

$$I(X) = \mathrm{E}\left[\left(\frac{\psi_f(Z)+Z}{f(Z)}\right)^2\right],$$

with $\psi_f = (\log f)'$. In particular, if F is a random function of the form $F(x) = Yx$ for $Y > 0$ some random variable independent of Z, then simple conditioning shows that the above becomes

$$I(X) = \mathrm{E}\left[\frac{Z^2}{Y^2}\right] = \mathrm{E}\left[\frac{1}{Y^2}\right],$$

so that

$$d_{\mathrm{TV}}(\phi, q_X) \le \frac{1}{\sqrt{2}}\sqrt{\mathrm{E}\left[\frac{1}{Y^2}\right] - 1 + \mathrm{E}(Y^2-1)}$$

where q_X refers to the density of $X \stackrel{d}{=} YZ$. This last inequality is to be compared with [8, Lemma 4.1] and also [34].

A Bounds for the supremum norm

First note that, for $l_y(x) = \delta_{\{x=y\}}$, the solution $f^p_{l_y}(x)$ of the Stein equation (3.7) is of the form

$$\frac{1}{p(x)}\int_a^x (\delta_{\{z=y\}} - p(y))p(z)dz = \frac{p(y)(\mathbb{I}_{[y,b)}(x) - P(x))}{p(x)}.$$

For all densities q such that $f^p_{l_y}(x) \in \mathcal{F}(q)$, Theorem 3.1 applies and yields

$$\sup_{y\in S}|p(y)-q(y)| \le \sup_{y\in S} p(y)\sqrt{\mathrm{E}_q[(\mathbb{I}_{[y,b)}(X) - P(X))^2/(p(X))^2]}\sqrt{\mathcal{J}(p,q)},$$

where b is either 0 or $+\infty$. We now prove that

$$\sup_{y\in S} p(y)\sqrt{\mathrm{E}_q[(\mathbb{I}_{[y,b)}(X) - P(X))^2/(p(X))^2]} \le 1$$

for $p(x) = c\,e^{-d|x|^\alpha}$ and any density q satisfying the assumptions of the claim. To this end note that straightforward manipulations lead to

$$
\begin{aligned}
&\mathrm{E}_q[\left(\mathbb{I}_{[y,b)}(X) - P(X)\right)^2 / (p(X))^2] \\
&\qquad = \frac{1}{c^2}\int_a^b q(x) e^{2d|x|^\alpha} (\mathbb{I}_{[y,b)}(x) - P(x))^2 dx \\
&\qquad = \frac{1}{c^2}\int_a^y q(x) e^{2d|x|^\alpha} (P(x))^2 dx + \frac{1}{c^2}\int_y^b q(x) e^{2d|x|^\alpha} (1 - P(x))^2 dx \\
&\qquad \leq \frac{1}{c^2} e^{2d|y|^\alpha} (P(y))^2 \int_a^y q(x) dx + \frac{1}{c^2} e^{2d|y|^\alpha} (1 - P(y))^2 \int_y^b q(x) dx \\
&\qquad = \frac{1}{c^2} e^{2d|y|^\alpha} (P(y))^2 + \frac{1}{c^2} e^{2d|y|^\alpha} (1 - 2P(y)) \mathrm{P}_q(X \geq y),
\end{aligned}
$$

where the inequality is due to the fact that $e^{2d|x|^\alpha} P(x)$ (resp., $e^{2d|x|^\alpha}(1 - P(x))$) is monotone increasing (resp., decreasing) on (a, y) (resp., (y, b)); see the proof of Theorem 4.1. This again directly leads to

$$
\begin{aligned}
&\mathrm{E}_q[\left(\mathbb{I}_{[y,b)}(X) - P(X)\right)^2 / (p(X))^2] \\
&\leq \sup_{y\in(a,b)} \left(c e^{-d|y|^\alpha} \sqrt{\frac{1}{c^2} e^{2d|y|^\alpha} ((P(y))^2 + (1 - 2P(y)) \mathrm{P}_q(X \geq y)} \right) \\
&= \sup_{y\in(a,b)} \left(\sqrt{(P(y))^2 + (1 - 2P(y)) \mathrm{P}_q(X \geq y)} \right).
\end{aligned}
$$

This last expression is equal to 1.

References

[1] George Afendras, Nickos Papadatos, and Vasilis Papathanasiou, *An extended Stein-type covariance identity for the Pearson family with applications to lower variance bounds*, Bernoulli **17** (2011), no. 2, 507–529. MR 2787602

[2] A. D. Barbour, *Stein's method for diffusion approximations*, Probab. Theory Related Fields **84** (1990), no. 3, 297–322.

[3] A. D. Barbour and Louis H. Y. Chen, *An introduction to stein's method*, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., vol. 4, Singapore University Press, Singapore, 2005.

[4] A. D. Barbour, O. Johnson, I. Kontoyiannis, and M. Madiman, *Compound Poisson approximation via information functionals*, Electron. J. Probab. **15** (2010), 1344–1368. MR 2721049 (2011k:94068)

[5] Andrew R. Barron, *Entropy and the central limit theorem*, Ann. Probab. **14** (1986), no. 1, 336–342. MR 815975 (87h:60048)

[6] Lawrence D. Brown, *A proof of the central limit theorem motivated by the Cramér-Rao inequality*, Statistics and probability: essays in honor of C. R. Rao, North-Holland, Amsterdam, 1982, pp. 141–148. MR 659464 (83h:60021)

[7] T. Cacoullos and V. Papathanasiou, *Characterizations of distributions by variance bounds*, Statist. Probab. Lett. **7** (1989), no. 5, 351–356. MR 1001133 (91e:62030)

[8] Theophilos Cacoullos, Vasilis Papathanasiou, and Sergey A. Utev, *Variational inequalities with examples and an application to the central limit theorem*, Ann. Probab. **22** (1994), no. 3, 1607–1618. MR 1303658 (95j:60036)

[9] Sourav Chatterjee, Jason Fulman, and Adrian Roellin, *Exponential approximation by exchangeable pairs and spectral graph theory*, ALEA **8** (2011), 1–27.

[10] Sourav Chatterjee and Qi-Man Shao, *Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie-Weiss model*, Ann. Appl. Probab. **21** (2011), no. 2, 464–483. MR 2807964 (2012b:60102)

[11] Louis H. Y. Chen, Larry Goldstein, and Qi-Man Shao, *Normal approximation by Stein's method*, Probability and its Applications (New York), Springer, Heidelberg, 2011. MR 2732624 (2012b:60103)

[12] Louis H. Y. Chen and Qi-Man Shao, *Stein's method for normal approximation*, An introduction to Stein's method, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., vol. 4, Singapore Univ. Press, Singapore, 2005, pp. 1–59. MR 2235448

[13] Thomas Cover and J. Thomas, *Elements of information theory*, vol. Second Edition, Wiley & Sons, New York, 2006.

[14] Christian Döbler, *Stein's method of exchangeable pairs for absolutely continuous, univariate distributions with applications to the polya urn model*, Preprint, arXiv:1207.0533, 2012.

[15] Alison L. Gibbs and Francis Edward Su, *On choosing and bounding probability metrics*, International Statistical Review / Revue Internationale de Statistique **70** (2002), no. 3, pp. 419–435 (English).

[16] Larry Goldstein and Gesine Reinert, *Distributional transformations, orthogonal polynomials, and Stein characterizations*, J. Theoret. Probab. **18** (2005), no. 1, 237–260. MR 2132278 (2006b:60026)

[17] ———, *Stein's method and the beta distribution*, Preprint, arxiv:1207.1460, 2012.

[18] Friedrich Götze, *On the rate of convergence in the multivariate clt*, Ann. Probab. **19** (1991), no. 2, 724–739.

[19] Oliver Johnson, *Information theory and the central limit theorem*, Imperial College Press, London, 2004. MR 2109042 (2006m:60001)

[20] Oliver Johnson and Andrew Barron, *Fisher information inequalities and the central limit theorem*, Probab. Theory Related Fields **129** (2004), no. 3, 391–409. MR 2128239 (2006b:62018)

[21] Ioannis Kontoyiannis, Peter Harremoës, and Oliver Johnson, *Entropy and the law of small numbers.*, IEEE Trans. Info. Theory **51** (2005), 466–472.

[22] Christophe Ley and Yvik Swan, *Stein's density approach for discrete distributions with applications to information inequalities*, Preprint, arXiv:1211.3668, 2012.

[23] Ju. V. Linnik, *An information-theoretic proof of the central limit theorem with Lindeberg conditions*, Theor. Probability Appl. **4** (1959), 288–299. MR 0124081 (23 #A1398)

[24] H. M. Luk, *Stein's method for the gamma distribution and related statistical applications*, Ph.D. thesis, University of Southern California, 1994.

[25] Eduardo Mayer-Wolf, *The Cramér-Rao functional and limiting laws*, Ann. Probab. **18** (1990), no. 2, 840–850. MR 1055436 (91m:60008)

[26] Ivan Nourdin and Giovanni Peccati, *Stein's method meets Malliavin calculus: a short survey with new estimates*, Recent development in stochastic dynamics and stochastic analysis, Interdiscip. Math. Sci., vol. 8, World Sci. Publ., Hackensack, NJ, 2010, pp. 207–236. MR 2807823

[27] ———, *Normal approximations with malliavin calculus : from stein's method to universality*, Cambridge Tracts in Mathematics, Cambridge University Press, 2012.

[28] A. Picket, *Rates of convergence of* χ^2 *approximations via stein's method*, Ph.D. thesis, Lincoln College, University of Oxford, 2004.

[29] Adrian Roellin, *On the optimality of stein factors*, Probability Approximations and Beyond (2012), 61–72.

[30] Nathan Ross, *Fundamentals of Stein's method*, Probab. Surv. **8** (2011), 210–293. MR 2861132

[31] Igal Sason, *Entropy bounds for discrete random variables via coupling*, Preprint, arXiv:1209.5259, 2012.

[32] ———, *An information-theoretic perspective of the Poisson approximation via the Chen-Stein method*, Preprint, arXiv:1206.6811v3, 2012.

[33] Wim Schoutens, *Orthogonal polynomials in Stein's method*, J. Math. Anal. Appl. **253** (2001), no. 2, 515–531. MR 1808151 (2001k:62021)

[34] R. Shimizu, *Error bounds for asymptotic expansion of the scale mixtures of the normal distribution*, Ann. Inst. Statist. Math. **39** (1987), 611–622.

[35] Ryoichi Shimizu, *On Fisher's amount of information for location family.*, A Modern Course on Statistical Distributions in Scientific Work, Springer, 1975, pp. 305–312.

[36] Charles Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory (Berkeley, Calif.), Univ. California Press, 1972, pp. 583–602. MR 0402873 (53 #6687)

[37] ———, *Approximate computation of expectations*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7, Institute of Mathematical Statistics, Hayward, CA, 1986. MR 882007 (88j:60055)

[38] Charles Stein, Persi Diaconis, Susan Holmes, and Gesine Reinert, *Use of exchangeable pairs in the analysis of simulations*, Stein's method: expository lectures and applications (Persi Diaconis and Susan Holmes, eds.), IMS Lecture Notes Monogr. Ser, vol. 46, Beachwood, Ohio, USA: Institute of Mathematical Statistics, 2004, pp. 1–26.

[39] Charles M. Stein, *Estimation of the mean of a multivariate normal distribution*, Ann. Statist. **9** (1981), no. 6, 1135–1151. MR 630098 (83a:62080)