Bernoulli **26**(3), 2020, 2051–2081
https://doi.org/10.3150/19-BEJ1182

First-order covariance inequalities via Stein's method

MARIE ERNST[1], GESINE REINERT[2] and YVIK SWAN[3]

[1]*Université de Liège, Liège, Belgium. E-mail: m.ernst@uliege.be*
[2]*University of Oxford, Oxford, UK. E-mail: reinert@stats.ox.ac.uk*
[3]*Université libre de Bruxelles, Bruxelles, Belgium. E-mail: yvswan@ulb.ac.be*

We propose probabilistic representations for inverse Stein operators (i.e., solutions to Stein equations) under general conditions; in particular, we deduce new simple expressions for the Stein kernel. These representations allow to deduce uniform and nonuniform Stein factors (i.e., bounds on solutions to Stein equations) and lead to new covariance identities expressing the covariance between arbitrary functionals of an arbitrary univariate target in terms of a weighted covariance of the derivatives of the functionals. Our weights are explicit, easily computable in most cases and expressed in terms of objects familiar within the context of Stein's method. Applications of the Cauchy–Schwarz inequality to these weighted covariance identities lead to sharp upper and lower covariance bounds and, in particular, weighted Poincaré inequalities. Many examples are given and, in particular, classical variance bounds due to Klaassen, Brascamp and Lieb or Otto and Menz are corollaries. Connections with more recent literature are also detailed.

Keywords: covariance identities; Cramér–Rao inequality; Stein equation; Stein kernel; variance bounds

1. Introduction

Much attention has been given in the literature to the problem of providing sharp tractable estimates on the variance of functions of random variables. Such estimates are directly related to fundamental considerations of pure mathematics (e.g., isoperimetric, logarithmic Sobolev and Poincaré inequalities), as well as essential issues from statistics (e.g., Cramér–Rao bounds, efficiency and asymptotic relative efficiency computations, maximum correlation coefficients and concentration inequalities).

One of the starting points of this line of research is Chernoff's famous result from [32] which states that, if $N \sim \mathcal{N}(0, 1)$, then

$$\mathbb{E}\big[g'(N)\big]^2 \le \mathrm{Var}\big[g(N)\big] \le \mathbb{E}\big[g'(N)^2\big] \tag{1}$$

for all sufficiently regular functions $g : \mathbb{R} \to \mathbb{R}$. Chernoff obtained the upper bound by exploiting orthogonality properties of the family of Hermite polynomials. The upper bound in (1) is, in fact, already available in [63] and is also a special case of the central inequality in [15]; see below. Cacoullos [16] extends Chernoff's bound to a wide class of univariate distributions (including discrete distributions) by proving that if $X \sim p$ has a density function p with respect to the Lebesgue measure then

$$\frac{\mathbb{E}[\tau_p(X)g'(X)]^2}{\mathrm{Var}[X]} \le \mathrm{Var}\big[g(X)\big] \le \mathbb{E}\big[\tau_p(X)g'(X)^2\big] \tag{2}$$

1350-7265

with $\tau_p(x) = p(x)^{-1} \int_x^{\infty} (t - \mathbb{E}[X]) p(t)\, \mathrm{d}t$. It is easy to see that, if p is the standard normal density, then $\tau_p(x) = 1$ so that (2) contains (1). Cacoullos also obtains a similar bound as (2) for discrete distributions on the positive integers, where the derivative is replaced by the forward difference and the weight becomes $\tau_p(x) = p(x)^{-1} \sum_{t=x+1}^{\infty} t p(t)$.

Variance inequalities such as (2) are closely related to a *Brascamp–Lieb inequality* from [15] which, in dimension 1, states that if $X \sim p$ and p is strictly log-concave then

$$\mathrm{Var}\big[g(X)\big] \le \mathbb{E}\left[\frac{(g'(X))^2}{(-\log p)''(X)}\right] \tag{3}$$

for all sufficiently regular functions g. In fact, the upper bound from (1) is an immediate consequence of (3) because, if p is the standard Gaussian density, then $(-\log p)''(x) \equiv 1$. This Brascamp–Lieb inequality is proved in [62] to be a consequence of Hoeffding's classical covariance inequality from [50], which states that if (X, Y) is a continuous bivariate random vector with cumulative distribution $H(x, y)$ and marginal cdfs $F(x)$, $G(x)$ then

$$\mathrm{Cov}\big[f(X), g(Y)\big] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f'(x)\big(H(x, y) - F(x)G(y)\big) g'(y)\, \mathrm{d}x\, \mathrm{d}y \tag{4}$$

under weak assumptions on f, g (see, e.g., [34]). The freedom of choice in the test functions f, g in (4) is exploited by [62] to prove that, if X has a C^2 strictly convex absolutely continuous density p then an *asymmetric Brascamp–Lieb inequality* holds:

$$\big|\mathrm{Cov}\big[f(X), g(X)\big]\big| \le \sup_x \left\{ \frac{|f'(x)|}{(\log p)''(x)} \right\} \mathbb{E}\big[\big|g'(X)\big|\big]. \tag{5}$$

In [23], identity (4) and inequalities (3) and (5) are extended to the multivariate setting and connections with logarithmic Sobolev inequalities for spin systems and related inequalities for log-concave densities are given. This material is extended in [71–73], providing applications in the context of isoperimetric inequalities and weighted Poincaré inequalities. In [34], the identity (4) is proved in all generality and used to provide expansions for the covariance in terms of canonical correlations and variables.

Further generalizations of Chernoff's bounds are provided in [18,19,29] and [51]; see also [14, 20,22,53,65] for the connection with probabilistic characterizations, as well as [43,77] for several generalizations in particular to stable distributions. Often by exploiting properties of suitable families of orthogonal polynomials, similar inequalities were obtained for univariate functionals of some specific multivariate distributions, for example, in [5,17,21,24,55,56]. A historical overview as well as a description of the connection between such bounds, the so-called Stein identities from Stein's method (see below) and Sturm–Liouville theory (see Section 4) can be found in [35]. Finally, we mention that all this material is closely connected to the study of the so-called *spectral gap* of the operator $\mathcal{L}f = f'' + (\log p)' f'$ which, in one dimension, is defined as

$$\lambda = \inf_g \frac{\mathbb{E}[(g'(X))^2]}{\mathrm{Var}[g(X)]},$$

where the infimum is taken over all functions $g \in C_0^\infty(\mathbb{R})$ such that $\mathrm{Var}[g(X)] > 0$. We will return to this briefly in Section 4.3, and refer the reader to [11,13,70] and [12] for an up-to-date overview of this topic.

To the best of our knowledge, the most general version of (1) and (2) is due to [52], where the following result is proved.

Theorem 1.1 (Klaassen bounds). *Let μ be some σ-finite measure. Let $\rho(x, y)$ be a measurable function such that $\rho(x, \cdot)$ does not change sign for μ almost $x \in \mathbb{R}$. Suppose that g is a measurable function such that $G(x) = \int \rho(x, y)g(y)\mu(\mathrm{d}y) + c$ is well-defined for some $c \in \mathbb{R}$. Let X be a real random variable with density p with respect to μ.*

- (*Klaassen upper variance bound*) *For all nonnegative measurable functions $h : \mathbb{R} \to \mathbb{R}$ such that $\mu(\{x \in \mathbb{R} | g(x) \neq 0, p(x)h(x) = 0\}) = 0$ we have*
$$\mathrm{Var}\big[G(X)\big] \leq \mathbb{E}\left[\frac{g(X)^2}{h(X)}\left(\frac{1}{p(X)}\int \rho(z, X)H(z)p(z)\mu(\mathrm{d}z)\right)\right] \tag{6}$$
with $H : \mathbb{R} \to \mathbb{R}$ supposed well-defined by $H(x) = \int \rho(x, y)h(y)\mu(\mathrm{d}y)$.
- (*Cramér–Rao lower variance bound*) *For all measurable functions $k : \mathbb{R} \to \mathbb{R}$ such that $0 < \mathbb{E}[k^2(X)] < \infty$ and $\mathbb{E}[k(X)] = 0$ we have*
$$\mathrm{Var}\big[G(X)\big] \geq \frac{\mathbb{E}[g(X)K(X)]^2}{\mathrm{Var}[k(X)]}, \tag{7}$$
where $K(x) = p(x)^{-1}\int \rho(z, x)k(z)p(z)\mu(\mathrm{d}z)$. Equality in (7) holds if and only if G is linear in k, p-almost everywhere.

Klaassen's proof of Theorem 1.1 relies on little more than the Cauchy–Schwarz inequality and Fubini's theorem; it has a slightly magical aura as little or no heuristic or context is provided as to the best choices of test functions h, k and kernel ρ or even to the nature of the weights appearing in (6) and (7). To the best of our knowledge, all available first-order variance bounds from the literature can be obtained from either (6) or (7) by choosing the appropriate test functions h or k and the appropriate kernel ρ. For instance, the weights appearing in the upper bound (6) generalize the Stein kernel from Cacoullos' bound (2) both in the discrete and the continuous case. Indeed taking $H(x) = x$ when the distribution p is continuous, $h(x) = 1$ and the weight becomes $p(x)^{-1}\int \rho(z, x)zp(z)\,\mathrm{d}\mu(z)$ which is none other than $\tau_p(x)$. A similar argument holds as well in the discrete case. In the same way, taking $k(x) = x$ leads to $K(x) = \tau_p(x)$ in (7), and thus the lower bound in (2) follows as well. The freedom of choice in the function h allows for much flexibility in the quality of the weights; this fact seems to date somewhat underexploited. This is perhaps due to the rather obscure nature of Klaassen's weights, a topic which we shall be one of the collateral learnings of this paper. Indeed we shall provide a natural theoretical home for Klaassen's result, in the framework of Stein's method.

Several variations on Klaassen's theorem have already been obtained via techniques related to Stein's method; an introduction of these techniques can be found in Section 2. The gist of the approach can nevertheless be understood very simply in the standard normal case. Stein's classical identity states that $N \sim \mathcal{N}(0,1)$ if and only if

$$\mathbb{E}\big[Ng(N)\big] = \mathbb{E}\big[g'(N)\big] \quad \text{for all bounded, continuous } g \text{ such that } \mathbb{E}\big[|g'(N)|\big] < \infty. \tag{8}$$

By the Cauchy–Schwarz inequality, we immediately deduce that, for all appropriate g,

$$\mathbb{E}\big[g'(N)\big]^2 = \mathbb{E}\big[N\big(g(N) - \mathbb{E}\big[g(N)\big]\big)\big]^2 \leq \mathbb{E}\big[N^2\big]\mathbb{E}\big[\big(g(N) - \mathbb{E}\big[g(N)\big]\big)^2\big] \leq \mathrm{Var}\big[g(N)\big], \tag{9}$$

which gives the lower bound in (1). For the upper bound, still by the Cauchy–Schwarz inequality,

$$\mathrm{Var}\big[g(N)\big] \leq \mathbb{E}\bigg[\bigg(\int_0^N g'(x)\,\mathrm{d}x\bigg)^2\bigg] \leq \mathbb{E}\bigg[N\int_0^N \big(g'(x)\big)^2\,\mathrm{d}x\bigg] = \mathbb{E}\big[\big(g'(N)\big)^2\big] \tag{10}$$

where the last identity is a direct consequence of Stein's identity (8) applied to the function $g(x) = \int_0^x (g'(u))^2\,\mathrm{d}u$. This is the upper bound in (1). The idea behind this proof is due to Chen [28]. As is now well known (again, we refer the reader to Section 2 for references and details), Stein's identity (8) for the normal distribution can be extended to basically any univariate (and even multivariate) distribution via a family of objects called "Stein operators". This leads to a wide variety of Stein-type integration by parts identities and it is natural to wonder whether Chen's approach can be used to obtain generalizations of Klaassen's theorem. First steps in this direction are detailed in [59,60]; in particular, it is seen that general lower variance bounds are easy to obtain from generalized Stein identities in the same way as in (9). Nevertheless, the method of proof in (10) for the upper bound cannot be generalized to arbitrary targets and, even in cases where the method does apply, the assumptions under which the bounds hold are quite stringent. To the best of our knowledge, the first to obtain upper variance bounds via properties of Stein operators is due to Saumard [71], by combining generalized Stein identities – expressed in terms of the Stein kernel $\tau_p(x)$ – with Hoeffding's identity (4). The scope of Saumard's weighted Poincaré inequalities is, nevertheless, limited and a general result such as Klaassen's is, to this date, not available in the literature.

There are obviously many applications of such material, not only towards considerations of pure mathematics but also to more applied questions, such as – in no particular order – questions from sensitivity analysis [70], stochastic ordering [66], the study of spin systems [62] and efficiency considerations [4]. We refer to all the above mentioned references for more references and details.

The main contributions of this paper can be categorized in two types:

1. *Covariance identities and inequalities.* The first main contribution of this paper is a generalization of Klaassen's variance bounds from Theorem 1.1 to covariance inequalities of arbitrary functionals of arbitrary univariate targets under minimal assumptions (see Theorems 3.1 and 3.5). Our results hereby therefore also contain basically the entire literature on the topic, in a unified framework containing in particular both continuous and discrete distributions alike. Moreover, the weights that appear in our bounds bear a clear and natural interpretation in terms of Stein operators which allow for easy computation for a wide variety of targets, as illustrated in the different examples we tackle as well as in Tables 1, 2 and 3 in the supplementary material [38]

in which we provide explicit variance bounds for univariate target distributions belonging to the classical integrated Pearson and Ord families (see Example 3.7 for a definition). In particular, Klaassen's bounds, its aforementioned corollaries as well as an (asymetric) Brascamp–Lieb and the weighted Poincaré inequality arise naturally in our setting. Moreover, in all these cases we recover freedom of choice in the weights which allows to weaken the underlying assumptions and extend the scope (e.g., to nonabsolutely continuous distributions).

2. *Stein operators and their properties.* The second main contribution of the paper lies in our method of proof, which contributes to the theory of Stein operators themselves. Specifically, we obtain several new probabilistic representations of inverse Stein operators (a.k.a. solutions to Stein equations) which open the way to a wealth of new manipulations which were hitherto unavailable. These representations also lead to new interpretations and ultimately new handles on several quantities which are crucial to the theory surrounding Stein's method (such as Stein kernels, Stein equations, Stein factors and Stein bounds). Finally, the various objects we identify provide natural connections with other topics of interest, and pave the way to several new research questions which we only briefly outline, for space reasons, in the final subsection of Section 4.

The paper is organized as follows. Section 2 contains the theoretical foundations of the paper. In Section 2.1 we recall the theory of canonical and standardized Stein operators introduced in [58] and introduce a (new) notion of inverse Stein operator (Definition 2.5). We also identify minimal conditions under which Stein-type probabilistic integration by parts formulas hold (see Lemmas 2.4 and 2.18). In Section 2.2, we provide the representation formulas for the inverse Stein operator (Lemmas 2.20 and 2.21). In Section 2.3, we clarify the conditions on the test functions under which the different identities hold, and provide bridges with the classical assumptions in the literature. Section 2.4 gives bounds on the solutions to the Stein equations. Section 3 contains the covariance identities and inequalities. After re-interpreting Hoeffding's identity (4) we obtain general and flexible lower and upper *covariance* bounds (Proposition 3.1 and Theorem 3.5). We then deduce Klaassen's bounds (Corollary 3.6) and provide examples for several concrete distributions, with more examples deferred to the three tables in the supplementary material [38] mentioned above. Finally, a discussion is provided in Section 4, wherein several examples are treated and connections with other theories are established, for instance, a Brascamp–Lieb inequality (Corollary 4.1) and Menz and Otto's asymmetric Brascamp–Lieb inequality (Corollary 4.3), as well as the link with an eigenfunction problem which can be seen as an extended Sturm–Liouville problem. For space reasons, technical proofs from Section 2, as well as tables and figures are provided in the supplementary material [38].

2. Stein differentiation

Stein's method consists in a collection of techniques for distributional approximation that was originally developed for normal approximation in [75] and for Poisson approximation in [27]; for expositions see, for example, the books [8,30,64,76] and the review papers [25,69]. Outside the Gaussian and Poisson frameworks, there exist several nonequivalent general theories allowing to setup Stein's method for many probability distributions, of which we single out the papers

[26,31,36,78,79] for univariate distributions under analytical assumptions, [6,7] for infinitely divisible distributions, [9] for discrete multivariate distributions and [40,46,47,61] for multivariate densities under diffusive assumptions.

The backbone of the present paper consists in the approach from [57,58,68]. Before introducing these results, we fix the notation. Let $\mathcal{X} \subset \mathbb{R}$ and equip it with some σ-algebra $\mathcal{A}$ and σ-finite measure μ. Let X be a random variable on $\mathcal{X}$, with probability measure $\mathbb{P}^X$ which is absolutely continuous with respect to μ; denote by p the corresponding probability density, and its support by $\mathcal{S}(p) = \{x \in \mathcal{X} : p(x) > 0\}$. As usual, $L^1(p)$ is the collection of all real valued functions f such that $\mathbb{E}|f(X)| < \infty$. We sometimes call the expectation under p the p-mean. In order to make the paper more concrete and readable, we shall exclude point mass distributions (Stein's method for point mass is available in [67]) and restrict our attention to distributions satisfying the following assumption.

Assumption A. The measure μ is either the counting measure on $\mathcal{X} = \mathbb{Z}$ or the Lebesgue measure on $\mathcal{X} = \mathbb{R}$. If μ is the counting measure, then there exist $a < b \in \mathbb{Z} \cup \{-\infty, \infty\}$ such that $\mathcal{S}(p) = [a, b] \cap \mathbb{Z}$. If μ is the Lebesgue measure, then there exist $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ such that $\mathcal{S}(p)^{\circ} =]a, b[$ and $\overline{\mathcal{S}(p)} = [a, b]$. Moreover, the measure μ is not point mass.

Let $\ell \in \{-1, 0, 1\}$. In the sequel, we shall restrict our attention to the following three derivative-type operators:

$$\Delta^{\ell} f(x) = \begin{cases} f'(x), & \text{if } \ell = 0; \\ \big(f(x+\ell) - f(x)\big)/\ell & \text{if } \ell \in \{-1, +1\}, \end{cases}$$

with $f'(x)$ the weak derivative defined Lebesgue almost everywhere, $\Delta^{+1} (\equiv \Delta^{+})$ the classical forward difference and $\Delta^{-1} (\equiv \Delta^{-})$ the classical backward difference. Whenever $\ell = 0$, we take μ as the Lebesgue measure and speak of the *continuous case*; whenever $\ell \in \{-1, 1\}$, we take μ as the counting measure and speak of the *discrete case*. There are two choices of derivatives in the discrete case, only one in the continuous case. We let $\mathrm{dom}(\Delta^{\ell})$ denote the collection of functions $f : \mathbb{R} \to \mathbb{R}$ such that $\Delta^{\ell} f(x)$ exists and is finite μ-almost surely. In the case $\ell = 0$, this corresponds to all absolutely continuous functions; in the case $\ell = \pm 1$ the domain is the collection of all functions on $\mathbb{Z}$. If $f \in \mathrm{dom}(\Delta^{\ell})$ is such that $\Delta^{\ell} f \mathbb{I}[a, b] \in L^1(\mu)$ then, for all c, d such that $a \le c \le d \le b$

$$\int_c^d \Delta^{\ell} f(x) \mu(\mathrm{d}x) = f(d + a_{\ell}) - f(c - b_{\ell}) \tag{11}$$

where

$$a_{\ell} = \mathbb{I}[\ell = 1] \quad \text{and} \quad b_{\ell} = \mathbb{I}[\ell = -1]. \tag{12}$$

We stress the fact that the values at c, d are understood as limits if either is infinite.

2.1. Stein operators and Stein equations

Our first definitions come from [58]. We first define $\mathrm{dom}(p, \Delta^\ell)$ as the collection of $f : \mathbb{R} \to \mathbb{R}$ such that $fp \in \mathrm{dom}(\Delta^\ell)$.

Definition 2.1 (Canonical Stein operators). Let $f \in \mathrm{dom}(p, \Delta^\ell)$ and consider the linear operator $f \mapsto \mathcal{T}_p^\ell f$ defined as $\mathcal{T}_p^\ell f(x) = p(x)^{-1}\Delta^\ell(f(x)p(x))p(x)$ for all $x \in \mathcal{S}(p)$ and as $\mathcal{T}_p^\ell f(x) = 0$ for $x \notin \mathcal{S}(p)$. The cases $\ell = 1$ and $\ell = -1$ provide the forward and backward Stein operators, denoted by $\mathcal{T}_p^+$ and $\mathcal{T}_p^-$, respectively; the case $\ell = 0$ provides the differential Stein operator denoted by $\mathcal{T}_p$. The operator $\mathcal{T}_p^\ell$ is called the canonical (ℓ-)Stein operator of p.

To describe the domain and the range of $\mathcal{T}_p^\ell$, we introduce the following sets of functions:

$$\mathcal{F}^{(0)}(p) = \{f \in L^1(p) : \mathbb{E}[f(X)] = 0\};$$

$$\mathcal{F}_\ell^{(1)}(p) = \Big\{ f \in \mathrm{dom}(p, \Delta^\ell) : \Delta^\ell(fp)\mathbb{I}[\mathcal{S}(p)] \in L^1(\mu) \text{ and } \int_{\mathcal{S}(p)} \Delta^\ell(fp)(x)\mu(\mathrm{d}x) = \mathbb{E}[\mathcal{T}_p^\ell f(X)] = 0 \Big\}.$$

We draw the reader's attention to the fact that the second condition in the definition of $\mathcal{F}_\ell^{(1)}(p)$ can be rewritten as $f(b + a_\ell)p(b + a_\ell) = f(a - b_\ell)p(a - b_\ell)$ combined example.

Example 2.2. If $p(x) = \phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ is the standard Gaussian density, then $\ell = 0$ and $\mathcal{T}_p^0 f(x) = f'(x) - xf(x)$ is the classical operator for this distribution. The set $\mathcal{F}_0^{(1)}(\phi)$ consists of all functions such $f'(x) - xf(x) \in L^1(\phi)$ and $\lim_{x\to-\infty} f(x)\phi(x) = \lim_{x\to\infty} f(x)\phi(x)$.

If $p(x) = p_\lambda(x) = e^{-\lambda}\lambda^x/x!$ is the Poisson density on $\mathbb{N}$, then $\ell \in \{-1, 1\}$; $\mathcal{T}_p^1 f(x) = \lambda f(x + 1)/(x + 1) - f(x)$ for all $x \in \mathbb{N}$ and 0 otherwise, and $\mathcal{T}_p^{-1} f(x) = f(x) - xf(x - 1)/\lambda$ for all $x \in \mathbb{N}$ and 0 otherwise. Both are equivalent, up to scaling, to the classical operator $\lambda f(x + 1) - xf(x)$ from [27] for this distribution. The set $\mathcal{F}_1^{(1)}(p_\lambda)$ consists of all functions such that $\lambda f(x + 1)/(x + 1) - f(x) \in L^1(p_\lambda)$ and $\lim_{x\to\infty} f(x)p_\lambda(x) = f(0)p_\lambda(0)$. Similarly, $\mathcal{F}_{-1}^{(1)}(p_\lambda)$ consists of all functions such that $f(x) - xf(x - 1)/\lambda \in L^1(p_\lambda)$ and $\lim_{x\to\infty} f(x)p_\lambda(x) = 0$.

The next lemma, which follows immediately from the definition of $\mathcal{T}_p^\ell f$ and of the different sets of functions, shows why $\mathcal{F}_\ell^{(1)}(p)$ is called the *canonical Stein class*.

Lemma 2.3 (Canonical Stein class). *For* $f \in \mathcal{F}_\ell^{(1)}(p)$, $\mathcal{T}_p^\ell f \in \mathcal{F}^{(0)}(p)$.

Crucially for the results in this paper, for all $f \in \mathrm{dom}(\Delta^\ell)$, $g \in \mathrm{dom}(\Delta^{-\ell})$ such that $f(\cdot) \times g(\cdot - \ell) \in \mathrm{dom}(\Delta^\ell)$ the operators Δ^ℓ satisfy the product rule

$$\Delta^\ell\big(f(x)g(x - \ell)\big) = \big(\Delta^\ell f(x)\big)g(x) + f(x)\Delta^{-\ell}g(x) \tag{13}$$

for all $\ell \in \{-1, 0, 1\}$. This product rule leads to an integration by parts (IBP) formula (a.k.a. Abel-type summation formula) as follows.

Lemma 2.4 (Stein IBP formula – version 1). *For all* $f \in \mathrm{dom}(p, \Delta^{\ell})$, $g \in \mathrm{dom}(\Delta^{-\ell})$ *such that* (i) $f(\cdot)g(\cdot - \ell) \in \mathcal{F}_{\ell}^{(1)}(p)$ *and* (ii) $f(\cdot)\Delta^{-\ell}g(\cdot) \in L^1(p)$, *we have*

$$\mathbb{E}\big[\big(\mathcal{T}_p^{\ell} f(X)\big)g(X)\big] = -\mathbb{E}\big[f(X)\Delta^{-\ell}g(X)\big]. \tag{14}$$

Proof. Under the stated assumptions, we can apply (13) to get

$$\mathcal{T}_p^{\ell}\big(f(x)g(x-\ell)\big) = \big(\mathcal{T}_p^{\ell} f(x)\big)g(x) + f(x)\big(\Delta^{-\ell}g(x)\big) \tag{15}$$

for all $x \in \mathcal{S}(p)$. Condition (i) in the statement guarantees that the left-hand side (l.h.s.) of (15) has mean 0, while condition (ii) guarantees that we can separate the expectation of the sum on the right-hand side (r.h.s.) into the sum of the individual expectations. □

A natural interpretation of (14) is that operator $\mathcal{T}_p^{\ell}$ is, in some sense to be made precise, the *skew-adjoint* operator to $\Delta^{-\ell}$ with respect to the scalar product $\langle f, g\rangle = \mathbb{E}[f(X)g(X)]$; this provides a supplementary justification to the use of the terminology "canonical" for operator $\mathcal{T}_p^{\ell}$. We discuss a consequence of this interpretation in Section 4. The conditions under which Lemma 2.4 holds are all but transparent. We clarify these assumptions in Section 2.3. For more details on Stein class and operators, we refer to [58] for the construction in an abstract setting, [57] for the construction in the continuous setting (i.e., $\ell = 0$) and [39] for the construction in the discrete setting (i.e., $\ell \in \{-1, 1\}$). Multivariate extensions are developed in [68].

The fundamental stepping stone for our theory is an inverse of the canonical operator $\mathcal{T}_p^{\ell}$ provided in the next definition.

Definition 2.5 (Canonical pseudo inverse Stein operator). Let $\ell \in \{-1, 0, 1\}$ and recall the notation a_{ℓ}, b_{ℓ} from (12). The canonical pseudo-inverse Stein operator $\mathcal{L}_p^{\ell}$ for the operator $\mathcal{T}_p^{\ell}$ is defined, for $h \in L^1(p)$, as

$$\begin{aligned}\mathcal{L}_p^{\ell} h(x) &= \frac{1}{p(x)} \int_a^{x-a_{\ell}} \big(h(u) - \mathbb{E}\big[h(X)\big]\big)p(u)\mu(\mathrm{d}u)\\ &= \frac{1}{p(x)} \int_{x+b_{\ell}}^{b} \big(\mathbb{E}\big[h(X)\big] - h(u)\big)p(u)\mu(\mathrm{d}u)\end{aligned} \tag{16}$$

for all $x \in \mathcal{S}(p)$, and $\mathcal{L}_p^{\ell} h(x) = 0$ for all $x \notin \mathcal{S}(p)$.

Equality between the second and third expressions in (16) is justified because $h \in L^1(p)$ so that the integral of $h(\cdot) - \mathbb{E}[h(X)]$ over the whole support cancels out. For example,

$$\mathcal{L}_p^{-} h(x) = \frac{1}{p(x)} \sum_{j=a}^{x} \big(h(j) - \mathbb{E}\big[h(X)\big]\big)p(j) = \frac{1}{p(x)} \sum_{j=x+1}^{b} \big(\mathbb{E}\big[h(X)\big] - h(j)\big)p(j).$$

Note that $\mathcal{L}_p^- h(b) = 0$ but $\mathcal{L}_p^- h(a) = h(a) - \mathbb{E}[h(X)]$. The denomination *pseudo-inverse-Stein operator* for $\mathcal{L}_p^\ell$ is justified by the following lemma whose proof is immediate.

Lemma 2.6. *For any $h \in L^1(p)$, $\mathcal{L}_p^\ell h \in \mathcal{F}_\ell^{(1)}(p)$. Moreover,* (i) *for all $h \in L^1(p)$ we have $\mathcal{T}_p^\ell \mathcal{L}_p^\ell h(x) = h(x) - \mathbb{E}[h(X)]$ at all $x \in \mathcal{S}(p)$ and* (ii) *for all $f \in \mathcal{F}_\ell^{(1)}(p)$ we have $\mathcal{L}_p^\ell \mathcal{T}_p^\ell f(x) = f(x) - f(a^+ - b_\ell)p(a^+ - b_\ell)/p(x) = f(x) - f(b^- + a_\ell)p(b^- + a_\ell)/p(x)$ at all $x \in \mathcal{S}(p)$. The operator $\mathcal{L}_p^\ell$ is invertible* (*with inverse $\mathcal{T}_p^\ell$*) *on the subclass of functions in $\mathcal{F}^{(0)}(p) \cap \mathcal{F}^{(1)}(p)$ which, moreover, satisfy $f(b^- + a_\ell)p(b^- + a_\ell) = f(a^+ - b_\ell)p(a^+ - b_\ell) = 0$.*

Starting from (15), we postulate the next definition.

Definition 2.7 (Standardizations of the canonical operator). Fix $\ell \in \{-1, 0, 1\}$ and $\eta \in L^1(p)$. The η-standardized Stein operator is

$$\mathcal{A}_p^{\ell,\eta} g(x) = \mathcal{T}_p^\ell \big(\mathcal{L}_p^\ell \eta(\cdot) g(\cdot - \ell)\big)(x) = \big(\eta(x) - \mathbb{E}\big[\eta(X)\big]\big) g(x) + \mathcal{L}_p^\ell \eta(x) \big(\Delta^{-\ell} g(x)\big) \quad (17)$$

acting on the collection $\mathcal{F}(\mathcal{A}_p^{\ell,\eta})$ of test functions g such that $\mathcal{L}_p^\ell \eta(\cdot) g(\cdot - \ell) \in \mathcal{F}_\ell^{(1)}(p)$ and $(\mathcal{L}_p^\ell \eta) \Delta^{-\ell} g \in L^1(p)$.

Example 2.8. If $p = \phi$ is the standard normal density and $\eta(x) = x$, then $\mathcal{L}_\phi^0 \eta(x) = -1$. More generally, if $\eta(x) = H_k(x) = (-1)^k e^{x^2/2} (d^k/dx^k) e^{-x^2/2}$ is the kth Hermite polynomial so that $H_{k+1}(x) = x H_k(x) - H_k'(x)$ then $\mathcal{L}_\eta^0 \eta(x) = -H_{k-1}(x)$. This leads to the family of standardized Stein operators $\mathcal{A}_\phi^{0,k} g(x) = H_k(x) g(x) - H_{k-1}(x) g'(x)$ already considered, for example, in [44].

If $p = p_\lambda$ and $\eta(x) = x$, then $\mathcal{L}_p^+ \eta(x) = -x$ and $\mathcal{L}_p^- \eta(x) = -\lambda$. This leads to the standardized operators $\mathcal{A}^+ g(x) = (x - \lambda) g(x) - x \Delta^- g(x) = \lambda g(x) - x g(x-1)$ and $\mathcal{A}^- g(x) = (x - \lambda) g(x) - \lambda \Delta^+ g(x) = -\lambda g(x+1) + x g(x)$; both are equivalent to the classical operator $\lambda g(x+1) - x g(x)$ first identified by Chen in [27]. Similarly, as for the Gaussian one could introduce the appropriate family of orthogonal polynomials (here the Charlier polynomials) and propose an entire family of operators; we refer to [44] for an overview.

Remark 2.9. The conditions appearing in the definition of $\mathcal{F}(\mathcal{A}_p^{\ell,\eta})$ are tailored to ensure that all identities and manipulations follow immediately. For instance, the requirement that $\mathcal{L}_p^\ell \eta(\cdot) g(\cdot - \ell) \in \mathcal{F}_\ell^{(1)}(p)$ in the definition of $\mathcal{F}(\mathcal{A}_p^{\ell,\eta})$ guarantees that the resulting functions $\mathcal{A}_p^{\ell,\eta} g(x)$ have p-mean 0 and the condition $(\mathcal{L}_p^\ell \eta) \Delta^{-\ell} g \in L^1(p)$ guarantees that the expectations of the individual summands on the r.h.s. of (17) exist. Again, our assumptions are not transparent; we discuss them in detail in Section 2.3.

The final ingredient for Stein differentiation is the *Stein equation*:

Definition 2.10 (Stein equation). Fix $\ell \in \{-1, 0, 1\}$ and $\eta \in L^1(p)$. For $h \in L^1(p)$, the $\mathcal{A}_p^{\ell,\eta}$-Stein equation for h is the equation $\mathcal{A}_p^{\ell,\eta} g(x) = h(x) - \mathbb{E}[h(X)]$, $x \in \mathcal{S}(p)$, that is,

$$\big(\eta(x) - \mathbb{E}\big[\eta(X)\big]\big)g(x) + \mathcal{L}_p^{\ell}\eta(x)\big(\Delta^{-\ell} g(x)\big) = h(x) - \mathbb{E}\big[h(X)\big], \quad x \in \mathcal{S}(p). \tag{18}$$

A solution to the Stein equation is any function $g \in \mathcal{F}(\mathcal{A}_p^{\ell,\eta})$ which satisfies (18) for all $x \in \mathcal{S}(p)$.

Our notation lead immediately to the next result.

Lemma 2.11 (Solution to the Stein equation). *Fix $\eta \in L^1(p)$. The Stein equation (18) for $h \in L^1(p)$ is solved by*

$$g_h^{p,\ell,\eta}(x) = \frac{\mathcal{L}_p^{\ell} h(x+\ell)}{\mathcal{L}_p^{\ell}\eta(x+\ell)} \tag{19}$$

with the convention that $g_h^{p,\ell,\eta}(x) = 0$ for all $x + \ell$ outside of $\mathcal{S}(p)$.

Proof. With $g = g_h^{p,\ell,\eta}$, by construction, $g \in \mathcal{F}(\mathcal{A}_p^{\ell,\eta})$, and

$$\begin{aligned}
\mathcal{A}_p^{\ell,\eta} g(x) &= \mathcal{T}_p^{\ell}\big(\mathcal{L}_p^{\ell}\eta(\cdot) g(\cdot - \ell)\big)(x) \\
&= \mathcal{T}_p^{\ell}\left(\mathcal{L}_p^{\ell}\eta(\cdot)\frac{\mathcal{L}_p^{\ell} h(\cdot)}{\mathcal{L}_p^{\ell}\eta(\cdot)}\right)(x) \\
&= \mathcal{T}_p^{\ell}\big(\mathcal{L}_p^{\ell} h(\cdot)\big)(x) \\
&= h(x) - \mathbb{E}\big[h(X)\big]
\end{aligned}$$

using Lemma 2.6 for the last step. Hence (18) is satisfied for all $x \in \mathcal{S}(p)$. □

When the context is clear, then we drop the superscripts and the subscript in g of (19). In the next two examples and beyond, the notation Id refers to the identity function $x \to \mathrm{Id}(x) = x$.

Example 2.12 (Binomial distribution). Let $0 < \theta < 1$ and $p(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ be the binomial density with parameters (n, θ) and $\mathcal{S}(p) = [0, n] \cap \mathbb{N}$. Stein's method for the binomial distribution was first developed in [37] using Δ^-. We fix $\eta(x) = x - n\theta$.

For $\ell = 1$, the class $\mathcal{F}_+^{(1)}(p)$ consists of functions $f : \mathbb{Z} \to \mathbb{R}$ which are bounded on $\mathcal{S}(p)$ and $f(0) = 0$, and $\mathcal{L}^+_{\mathrm{bin}(n,\theta)}\eta(x) = -(1-\theta)x$, leading to

$$\mathcal{A}^{+,\mathrm{Id}}_{\mathrm{bin}(n,\theta)} g(x) = (x - n\theta)g(x) - (1-\theta)x\Delta^- g(x) \tag{20}$$

with corresponding class $\mathcal{F}(\mathcal{A}^{+,\mathrm{Id}}_{\mathrm{bin}(n,\theta)})$ which contains all functions $g : \mathbb{Z} \to \mathbb{R}$. The solution to the $\mathcal{A}^{+,\mathrm{Id}}_{\mathrm{bin}(n,\theta)}$-Stein equation (see (18)) is $g^+(n) = 0$ and

$$g^+(x) = \frac{-1}{(1-\theta)(x+1)p(x+1)} \sum_{j=0}^{x} \big(h(j) - \mathbb{E}\big[h(X)\big]\big)p(j) \quad \text{for all } 0 \le x \le n-1.$$

For $\ell = -1$, the class $\mathcal{F}^{(1)}_{-}(p)$ consists of functions $f : \mathbb{Z} \to \mathbb{R}$ which are bounded on $\mathcal{S}(p)$ and such that $f(n) = 0$, and $\mathcal{L}^{-}_{\mathrm{bin}(n,\theta)}\eta(x) = -\theta(n-x)$, leading to

$$\mathcal{A}^{-,\mathrm{Id}}_{\mathrm{bin}(n,\theta)} g(x) = (x - n\theta)g(x) - \theta(n-x)\Delta^+ g(x)$$

on the same class as (20). The solution to the $\mathcal{A}^{-,\mathrm{Id}}_{\mathrm{bin}(n,\theta)}$-Stein equation is $g^-(0) = 0$ and

$$g^-(x) = \frac{-1}{\theta(n-(x-1))p(x-1)} \sum_{j=0}^{x-1} \big(h(j) - \mathbb{E}\big[h(X)\big]\big)p(j) \quad \text{for all } 1 \le x \le n.$$

The function $-g^-$ is studied in [37] where bounds on $\|\Delta^- g^-\|$ are provided (see equation (10) in that paper); see also Section 2.4 where bounds on $\|g^-\|$ are provided.

Example 2.13 (Beta distribution). Let $p(x) = x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha,\beta)$ be the beta density with parameters (α,β) and $\mathcal{S}(p) = (0,1)$. Stein's method for the beta distribution was developed in [36,45] using the Stein operator $\mathcal{A}f(x) = x(1-x)f'(x) + (\alpha(1-x) - \beta x)f(x)$. In our notation, we have $\ell = 0$ and $\mathcal{F}^{(1)}_0(p)$ consists of functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(0^+)p(0^+) = f(1^-)p(1^-)$ and $|(fp)'|$ is Lebesgue integrable on $[0,1]$. Fixing $\eta(x) = x - \alpha/(\alpha+\beta)$ gives $\mathcal{L}_{\mathrm{beta}(\alpha,\beta)}\eta(x) = -x(1-x)/(\alpha+\beta)$ leading to the operator

$$\mathcal{A}^{\mathrm{Id}}_{\mathrm{Beta}(\alpha,\beta)} g(x) = \left(x - \frac{\alpha}{\alpha+\beta}\right)g(x) - \frac{x(1-x)}{\alpha+\beta} g'(x)$$

with domain $\mathcal{F}(\mathcal{A}^{\mathrm{Id}}_{\mathrm{Beta}(\alpha,\beta)})$ the set of differentiable functions $g : \mathbb{R} \to \mathbb{R}$ such that $x(1-x)g(x) \in \mathcal{F}^{(1)}_0(p)$ and $x(1-x)g'(x) \in L^1(p)$. The solution to the $\mathcal{A}^{\mathrm{Id}}_{\mathrm{Beta}(\alpha,\beta)}$-Stein equation is

$$g(x) = \frac{-(\alpha+\beta)}{x(1-x)p(x)} \int_0^x \big(h(u) - \mathbb{E}\big[h(X)\big]\big)p(u)\,du, \quad x \in (0,1).$$

The operator $\mathcal{A}^{\mathrm{Id}}_{\mathrm{Beta}(\alpha,\beta)} f$ is, up to multiplication by $\alpha+\beta$, the classical Stein operator $\mathcal{A}f$ for the beta density; see [36,45] for details and bounds on solutions and their derivatives. See also Section 2.4 where bounds on $\|g\|$ are provided.

In order to propose a more general example, we recall the concept of a Stein kernel, here extended to continuous and discrete distributions alike.

Definition 2.14 (The Stein kernel). Let $X \sim p$ have finite mean. The (ℓ-)Stein kernel of X (or of p) is the function

$$\tau_p^\ell(x) = -\mathcal{L}_p^\ell(\mathrm{Id})(x).$$

Metonymously, we refer to the random variable $\tau_p^\ell(X)$ as the (ℓ-)Stein kernel of X.

Remark 2.15. The function $\tau_p^\ell(\cdot)$ is studied in detail for $\ell = 0$ in [76], Lecture VI. This function is particularly useful for Pearson (and discrete Pearson a.k.a. Ord) distributions which are characterized by the fact that their Stein kernel τ_p^ℓ is a second degree polynomial, see Example 3.7. For more on this topic, we also refer to [39] as well as [33,41,42] wherein important contributions to the theory of Stein kernels are provided in a multivariate setting.

The next example gives some (ℓ-)Stein kernels, exploiting the fact that if the mean of X is ν, then $\mathcal{L}_p^\ell(\mathrm{Id})(x) = \mathcal{L}_p^\ell(\mathrm{Id} - \nu)(x)$.

Example 2.16. If $X \sim \mathrm{Bin}(n, \theta)$, then using $\eta(x) = x - n\theta$, Example 2.12 gives $\tau^+_{\mathrm{bin}(n,\theta)}(x) = (1-\theta)x$ and $\tau^-_{\mathrm{bin}(n,\theta)}(x) = \theta(n-x)$. If $X \sim \mathrm{Beta}(\alpha, \beta)$ then Example 2.13 with $\eta(x) = x - \alpha/(\alpha+\beta)$ gives $\tau^0_{\mathrm{Beta}(\alpha,\beta)}(x) = x(1-x)/(\alpha+\beta)$.

Example 2.17 (A general example). Let p satisfy Assumption A and suppose that it has finite mean ν. Fixing $\eta(x) = x - \nu$, operator (17) becomes

$$\mathcal{A}_p^{\tau_p^\ell} g(x) = (x-\nu)g(x) - \tau_p^\ell(x)\Delta^{-\ell} g(x)$$

with corresponding class $\mathcal{F}(\mathcal{A}_p^{\tau_p^\ell})$ which contains all functions $g : \mathbb{R} \to \mathbb{R}$ such that $\tau_p^\ell(\cdot)g(\cdot - \ell) \in \mathcal{F}_\ell^{(1)}(p)$ and $\tau_p^\ell \Delta^{-\ell} g \in L^1(p)$. Again, we stress that such conditions are clarified in Section 2.3. Using Lemma 2.11, the solution to the $\mathcal{A}_p^{\tau_p}$ Stein equation is

$$g_{\mathrm{Id}}^{p,\ell,h}(x) = \frac{-\mathcal{L}_p^\ell h(x+\ell)}{\tau_p^\ell(x+\ell)}.$$

Bounds on $\|g\|$ are provided in Section 2.3. Stein's method based on $\mathcal{A}_p^{\tau_p^\ell}$ is already available in several important subcases, for example, in [36,54,74] for continuous distributions.

The construction is tailored to ensure that all operators have mean 0 over the entire classes of functions on which they are defined. We immediately deduce the following family of Stein integration by parts formulas.

Lemma 2.18 (Stein IBP formula – version 2). *Let $X \sim p$. Then*

$$\mathbb{E}\left[-\left\{\mathcal{L}_p^\ell f(X)\right\}\Delta^{-\ell} g(X)\right] = \mathbb{E}\left[\left(f(X) - \mathbb{E}\left[f(X)\right]\right)g(X)\right] \tag{21}$$

for all $f \in L^1(p)$, $g \in \mathrm{dom}(\Delta^{-\ell})$ *such that* $\mathcal{L}_p^\ell f(\cdot)g(\cdot - \ell) \in \mathcal{F}_\ell^{(1)}(p)$ *and* $\mathcal{L}_p^\ell f(\cdot)\Delta^{-\ell}g(\cdot) \in L^1(p)$.

Proof. Identity (21) follows directly from the Stein product rule in [58], Theorem 3.24, or by using the fact that expectations of the operators in (17) are equal to 0. □

For our future developments, it is important to note that in Lemma 2.18 the test functions f and g do not play a symmetric role. If $g \in L^1(p)$, then the right-hand side of (21) is the covariance $\mathrm{Cov}(f(X), g(X))$. Similarly, as for Lemma 2.4, the conditions under which Lemma 2.18 applies are not transparent in their present form. In Section 2.3, various explicit sets of conditions are provided under which the IBP (21) is applicable.

2.2. Representations of the inverse Stein operator

This section contains the first main results of the paper, namely probabilistic representations for this operator. We start with a simple rewriting of $\mathcal{L}_p^\ell$. Given $\ell \in \{-1, 0, 1\}$, recall the notation $a_\ell = \mathbb{I}[\ell = 1]$ and define

$$\chi^\ell(x, y) = \mathbb{I}[x \le y - a_\ell]. \tag{22}$$

Such generalized indicator functions particularize, in the three cases that interest us, to $\chi^0(x, y) = \mathbb{I}[x \le y]$ ($\ell = 0$), $\chi^-(x, y) = \mathbb{I}[x \le y]$ ($\ell = -1$) and $\chi^+(x, y) = \mathbb{I}[x < y]$ ($\ell = 1$). Their properties lead to some form of "calculus" as follows.

Lemma 2.19 (Chi calculation rules). *The function $\chi^\ell(x, y)$ is nonincreasing in x and nondecreasing in y. For all x, y, we have*

$$\chi^\ell(x, y) + \chi^{-\ell}(y, x) = 1 + \mathbb{I}[\ell = 0]\mathbb{I}[x = y] \quad \textit{and} \tag{23}$$

$$\begin{aligned} &\chi^\ell(u, y)\chi^\ell(v, y) = \chi^\ell\big(\max(u, v), y\big) \quad \textit{and} \\ &\chi^\ell(x, u)\chi^\ell(x, v) = \chi^\ell\big(x, \min(u, v)\big). \end{aligned} \tag{24}$$

Let p with support $\mathcal{S}(p)$ satisfy Assumption A. Then for any $f \in L^1(p)$, with (16),

$$\begin{aligned} \mathcal{L}_p^\ell f(x) &= \frac{1}{p(x)}\mathbb{E}\big[\chi^\ell(X, x)\big(f(X) - \mathbb{E}\big[f(X)\big]\big)\big] \\ &= \frac{1}{p(x)}\mathbb{E}\big[\big(\chi^\ell(X, x) - \mathbb{E}\big[\chi^\ell(X, x)\big]\big)\big(f(X) - \mathbb{E}\big[f(X)\big]\big)\big]. \end{aligned} \tag{25}$$

The function

$$\Phi_p^\ell(u, x, v) = \frac{\chi^\ell(u, x)\chi^{-\ell}(x, v)}{p(x)}, \quad x \in \mathcal{S}(p) \tag{26}$$

set to be 0 for $x \notin \mathcal{S}(p)$, is used in the following representation for the Stein inverse operator.

Lemma 2.20 (Representation formula I). *Let X, X_1, X_2 be independent copies of $X \sim p$ with support $\mathcal{S}(p)$. Then, for all $f \in L^1(p)$ we have*

$$-\mathcal{L}_p^\ell f(x) = \mathbb{E}\big[\big(f(X_2) - f(X_1)\big)\Phi_p^\ell(X_1, x, X_2)\big]. \tag{27}$$

Proof. The $L^1(p)$ condition on f suffices for the expectation on the r.h.s. of (27) to be finite for all $x \in \mathcal{S}(p)$. Suppose without loss of generality that $\mathbb{E}[f(X)] = 0$. Using that X_1, X_2 are i.i.d., we reap

$$\begin{aligned}
&\mathbb{E}\big[\big(f(X_2) - f(X_1)\big)\chi^\ell(X_1, x)\chi^{-\ell}(x, X_2)\big] \\
&\quad = \mathbb{E}\big[\chi^\ell(X_1, x)\big]\mathbb{E}\big[f(X_2)\chi^{-\ell}(x, X_2)\big] - \mathbb{E}\big[f(X_1)\chi^\ell(X_1, x)\big]\mathbb{E}\big[\chi^{-\ell}(x, X_2)\big] \\
&\quad = \mathbb{E}\big[\chi^\ell(X_1, x)\big]\mathbb{E}\big[f(X_2)\big(1 - \chi^\ell(X_2, x)\big)\big] - \mathbb{E}\big[f(X_1)\chi^\ell(X_1, x)\big]\mathbb{E}\big[\chi^{-\ell}(x, X_2)\big] \\
&\quad = -\mathbb{E}\big[f(X)\chi^\ell(X, x)\big]\big(\mathbb{E}\big[\chi^\ell(X, x)\big] + \mathbb{E}\big[\chi^{-\ell}(x, X)\big]\big),
\end{aligned}$$

where in the third line we used the fact that $\mathbb{E}[f(X)\mathbb{I}[\ell = 0]\mathbb{I}[X = x]] = 0$ under the stated assumptions. For the same reasons, we have $\mathbb{E}[\chi^\ell(X, x) + \chi^{-\ell}(x, X)] = 1$ for all $x \in \mathcal{X}$ and all $\ell \in \{-1, 0, 1\}$. The conclusion follows by recalling (25). □

The function defined in (26) allows to perform "probabilistic integration" as follows: if $f \in \mathrm{dom}(\Delta^{-\ell})$ is such that $(\Delta^{-\ell} f)$ is integrable on $[x_1, x_2] \cap \mathcal{S}(p)$ then

$$f(x_2) - f(x_1) = \mathbb{E}\big[\Phi_p^\ell(x_1, X, x_2)\Delta^{-\ell} f(X)\big] = \begin{cases} \displaystyle\int_{x_1}^{x_2} f'(u)\,\mathrm{d}u & (\ell = 0) \\ \displaystyle\sum_{j=x_1}^{x_2-1} \Delta^+ f(j) & (\ell = -1) \\ \displaystyle\sum_{j=x_1+1}^{x_2} \Delta^- f(j) & (\ell = 1) \end{cases} \tag{28}$$

for all $x_1 < x_2 \in \mathcal{S}(p)$. If, furthermore, $f \in L^1(p)$ then (by a conditioning argument)

$$\mathbb{E}\big[\big(f(X_2) - f(X_1)\big)\mathbb{I}[X_1 < X_2]\big] = \mathbb{E}\big[\Phi_p^\ell(X_1, X, X_2)\Delta^{-\ell} f(X)\big].$$

Equation (28) leads to the next representation formula for the inverse Stein operator.

Lemma 2.21 (Representation formula II). *Let $X \sim p$. Define the kernel K_p on $\mathcal{S}(p) \times \mathcal{S}(p)$ by*

$$K_p^\ell\big(x, x'\big) = \mathbb{E}\big[\chi^\ell(X, x)\chi^\ell\big(X, x'\big)\big] - \mathbb{E}\big[\chi^\ell(X, x)\big]\mathbb{E}\big[\chi^\ell\big(X, x'\big)\big]. \tag{29}$$

Then $K_p^\ell(x, x')$ is symmetric and positive. Moreover, for all $f \in \mathrm{dom}(\Delta^{-\ell}) \cap L^1(p)$,

$$-\mathcal{L}_p^\ell f(x) = \mathbb{E}\left[\frac{K_p^\ell(X, x)}{p(X)p(x)}\Delta^{-\ell} f(X)\right]. \tag{30}$$

Proof. Symmetry of K_p^ℓ is immediate. Applying first (24) and then (23),

$$K_p^\ell(x,x') = \mathbb{E}[\chi^\ell(X,\min(x,x'))](1-\mathbb{E}[\chi^\ell(X,\max(x,x'))]) \\ = \mathbb{E}[\chi^\ell(X,\min(x,x'))]\mathbb{E}[\chi^{-\ell}(\max(x,x'),X)] \tag{31}$$

which is necessarily positive. To prove (30), we insert (28) into (27), to obtain

$$-\mathcal{L}_p^\ell f(x) = \mathbb{E}[\Delta^{-\ell} f(X')\Phi_p^\ell(X_1,X',X_2)\Phi_p^\ell(X_1,x,X_2)] \\ = \mathbb{E}[\Delta^{-\ell} f(X')\mathbb{E}[\Phi_p^\ell(X_1,X',X_2)\Phi_p^\ell(X_1,x,X_2)|X']].$$

For all $x, x' \in \mathcal{S}(p)$, by (24),

$$\mathbb{E}[\Phi_p^\ell(X_1,x,X_2)\Phi_p^\ell(X_1,x',X_2)] \\ = \frac{1}{p(x)p(x')}\mathbb{E}[\chi^\ell(X,x)\chi^\ell(X,x')]\mathbb{E}[\chi^{-\ell}(x,X)\chi^{-\ell}(x',X)] \\ = \frac{1}{p(x)p(x')}(\mathbb{E}[\chi^\ell(X,\min(x,x'))]\mathbb{E}[\chi^{-\ell}(\max(x,x'),X)]).$$

Using (31), we recognize the kernel $K_p^\ell(x,x')$ in the numerator, and identity (30) follows. □

Example 2.22. Representations (27) and (30) can easily be applied to obtain representations for the Stein kernel $\tau_p^\ell(x)$:

$$\tau_p^\ell(x) = -\mathcal{L}_p^\ell(\mathrm{Id})(x) \\ = \mathbb{E}[(X_2-X_1)\Phi_p^\ell(X_1,x,X_2)] = \mathbb{E}\left[\frac{K_p^\ell(X,x)}{p(X)p(x)}\right].$$

In particular, the Stein kernel is positive on $\mathcal{S}(p)$.

Identity (27) seems to be new, although it is present in nonexplicit form in [26], equation (4.16). Representation (30) is, in the continuous $\ell = 0$ case, already available in [71]. The kernel $K_p^\ell(x,x')$ is a classical object in the theory of covariance representations and inequalities; an early appearance is attributed by [66] to [50] (see [49], pp. 57–109, for an English translation). The perhaps not very surprising extension to the discrete case is, to the best of our knowledge, new.

As a first result from our set-up, (30) applied to the function $f(x) = \mathcal{T}_p^\ell 1(x)$ immediately gives the following (see the supplementary material [38] for a proof).

Proposition 2.23 (Menz–Otto formula). *Suppose that the constant function* 1 *belongs to* $\mathcal{F}_\ell^{(1)}(p)$, *that* $-\Delta^{-\ell}\mathcal{T}_p^\ell 1(x) > 0$ *for almost all* $x \in \mathcal{S}(p)$ *and* $\Delta^{-\ell}(\mathcal{T}_p^\ell 1) \in L^1(\mu)$. *Then, for* *every $x' \in \mathcal{S}(p)$, the function*

$$p^{\star}_{x'}(x) = \frac{K_p^{\ell}(x, x')}{p(x')} \left(-\Delta^{-\ell} \mathcal{T}_p^{\ell} 1(x)\right) \tag{32}$$

is a density on $\mathcal{S}(p)$ with respect to μ.

Remark 2.24. If $K_p^{\ell}(x, x)/p(x)$ is bounded, then the assumptions in Proposition 2.23 are satisfied as soon as $-\Delta^{-\ell}\mathcal{T}_p^{\ell} 1 \in L^1(\mu)$. The proposition thus applies when $\ell = 0$ and $p(x) = e^{-H(x)}$ with H a strictly convex function such that $\lim_{x\to\pm\infty} H'(x) = 0$. This puts us in the context studied by [62] and formula (32) is equivalent to their [62], equation (14); we return to this in Section 3.

The next proposition gives some properties of $K_p^{\ell}(x, x')$.

Proposition 2.25. (i) $K_p^{\ell}(x, x') \le K_p^{\ell}(\min(x, x'), \min(x, x'))$ *for all* $x, x' \in \mathcal{S}(p)$. (ii) *If* $\mathbb{E}[\chi^{\ell}(X, x)]/p(x)$ *is nondecreasing, then the function* $x \mapsto K_p^{\ell}(x, x')/p(x)$ *is non-decreasing for* $x < x'$. (iii) *If* $\mathbb{E}[\chi^{-\ell}(x, X)]/p(x)$ *is nonincreasing, then the function* $x \mapsto K_p^{\ell}(x, x')/p(x)$ *is nonincreasing for* $x > x'$.

Figures 1 and 2 in the supplementary material [38] display the functions $x \mapsto K_p^{\ell}(x, x')/p(x)$ (for various values of x') and $x \mapsto K_p^{\ell}(x, x)/p(x)$ for the standard normal and several parameter choices in beta, gamma, binomial, Poisson and hypergeometric distributions.

Example 2.26. The following facts are easy to prove:

1. If $p(x)$ is the standard normal distribution, then $K_p^0(x, x)/p(x)$ behaves as $1/|x|$ for large $|x|$; see Figure 2(a).
2. If $p(x)$ is gamma, then $K_p^0(x, x)/p(x)$ behaves as a constant for large x; see Figure 2(c).
3. The function $x \mapsto K_p^0(x, x)/p(x)$ is not in $L^1(p)$ for p a Cauchy distribution.
4. If p is strictly-log concave, then $K_p^{\ell}(x, x)/p(x)$ is bounded.

2.3. Sufficient conditions and integrability

As anticipated, we now study the conditions under which the IBP Lemmas 2.4 and 2.18 hold. All proofs are technical manipulations of basic calculus and relegated to Section 1 in the supplementary material.

We start by the decryption of the conditions for Lemma 2.4. Recall the notation a_{ℓ} and b_{ℓ} from (12). Furthermore, if $\ell = 0$ we write $f(a^+) = \lim_{x\to a, x>a} f(x)$ and $f(b^-) = \lim_{x\to b, x<b} f(x)$. In the case that $a = -\infty$ or $b = \infty$, for $\ell \in \{-1, 0, 1\}$, we write $f(-\infty^+) = \lim_{x\to-\infty} f(x)$ and $f(\infty^-) = \lim_{x\to\infty} f(x)$. To simplify notation, if $\ell \in \{-1, 1\}$ and $a \ne -\infty$, we write $f(a^+) = f(a)$, and similarly, if $b \ne \infty$, $f(b^-) = f(b)$.

Proposition 2.27 (Sufficient conditions for IBP – version 1). *Let* $f \in \mathrm{dom}(p, \Delta^{\ell})$ *and* $g \in \mathrm{dom}(\Delta^{-\ell})$. *In order for* (14) *to hold, it suffices that they jointly satisfy*

$$\big(\mathcal{T}_p^{\ell} f\big) g \text{ and } f\big(\Delta^{-\ell} g\big) \in L^1(p) \tag{33}$$

$$f\big(b^- + a_\ell\big) g\big(b^- + a_\ell - \ell\big) p\big(b^- + a_\ell\big) = f\big(a^+ - b_\ell\big) g\big(a^+ - b_\ell - \ell\big) p\big(a^+ - b_\ell\big). \tag{34}$$

We now derive (almost) necessary and sufficient conditions for (21) to hold.

Proposition 2.28 (Sufficient conditions for IBP – version 2). *Let* $g \in \mathrm{dom}(\Delta^{-\ell})$. *In order for* (21) *to hold, it is necessary and sufficient that they jointly satisfy*

$$f, g \text{ and } fg \in L^1(p), \tag{35}$$

$$\mathcal{L}_p^{\ell} f\big(\Delta^{-\ell} g\big) \in L^1(p) \tag{36}$$

$$g\big(b^- + a_\ell - \ell\big) p\big(b^- + a_\ell\big) \mathcal{L}_p^{\ell} f\big(b^- + a_\ell\big) = g\big(a^+ - b_\ell - \ell\big) p\big(a^+ - b_\ell\big) \mathcal{L}_p^{\ell} f\big(a^+ - b_\ell\big). \tag{37}$$

Requirement (35) is natural and condition (37) is mild as it is satisfied as soon as g and/or f are well behaved at the edges of the support. We specialize condition (37) further in our next result.

Proposition 2.29. *Let* f, g *and* $fg \in L^1(p)$. *If* $g \in \mathrm{dom}(\Delta^{-\ell})$ *is of bounded variation and satisfies the following two conditions*:

1. $g(a^+ - b_\ell - \ell)\mathbb{P}(X \le a^+ - a_\ell - b_\ell) = 0$ *and* $g(b^- + a_\ell - \ell)\mathbb{P}(X \ge b^- + a_\ell + b_\ell) = 0$,
2. $g(a^+ - b_\ell - \ell)\mathbb{E}[|f(X)|\chi^{\ell}(X \le a^+ - b_\ell)] = 0$ *and* $g(b^- + a_\ell - \ell)\mathbb{E}[|f(X)|\chi^{-\ell}(b^- + a_\ell, X)] = 0$,

then (37) *holds. In particular, if* f *is bounded or in* $L^2(p)$, *then the condition* 2 *above is implied by Condition* 1.

Remark 2.30. The assumptions in Lemma 2.29 are related to those in [71] in the case $\ell = 0$. The main difference to these classical assumptions is that Lemma 2.29 only imposes conditions on one of the functions. We stress that there is a certain degree of redundancy in the items 1 and 2 together with the assumption that $g \in L^1(p)$ and is of bounded variation; the statement could be shortened at the loss of readability.

2.4. The inverse Stein operator

We conclude this section by exploring easy consequences of the representations from Section 2.2. These results are also of independent interest in Stein's method.

Lemma 2.31. *If $f, \mathcal{L}_p^\ell f(X) \in L^1(p)$, then*

$$\begin{aligned}\mathbb{E}\big[-\mathcal{L}_p^\ell f(X)\big] &= \mathbb{E}\big[(X_2 - X_1)^+\big(f(X_2) - f(X_1)\big)\big] \\ &= \frac{1}{2}\mathbb{E}\big[(X_2 - X_1)\big(f(X_2) - f(X_1)\big)\big] \qquad (38)\end{aligned}$$

where $(\cdot)^+$ denotes the positive part of $(\cdot)$. In particular, if the conditions of Lemma 2.18 are satisfied with $f(x) = g(x) = \mathrm{Id}$, then $\mathbb{E}[\tau_p^\ell(X)] = \mathrm{Var}(X)$.

Proof. Representation (27) gives $\mathbb{E}[-\mathcal{L}_p^\ell f(X)] = \mathbb{E}[(f(X_2) - f(X_1))\Phi_p^\ell(X_1, X, X_2)]$. Using (28) with $f(x) = x$, we have $\mathbb{E}[\Phi_p^\ell(x_1, X, x_2)] = (x_2 - x_1)^+$. Hence, after conditioning with respect to X_1, X_2, the first equality in (38) follows. The second equality follows by symmetry. The second claim is immediate under the stated assumptions. □

Remark 2.32. Once again, our assumptions are minimal but not transparent. It is easy to spell out these conditions explicitly for any specific target. For instance, if X has bounded support or support $\mathbb{R}$ then finite variance suffices.

Proposition 2.33. *Suppose that all test functions satisfy the conditions in Lemma 2.18. Let $\|f\|_{\mathcal{S}(p),\infty} = \sup_{x \in \mathcal{S}(p)} |f(x)|$.*

1. *If f is monotone, then $\mathcal{L}_p^\ell f(x)$ does not change sign.*
2. *(Uniform bounds Stein bounds) Consider, for h and η in $L^1(p)$ the function*

$$g_h^{p,\ell,\eta}(x) = \frac{\mathcal{L}_p^\ell h(x+\ell)}{\mathcal{L}_p^\ell \eta(x+\ell)}$$

defined in (19) which solves the η-Stein equation (18) for h. If η is monotone and $|h(x) - h(y)| \le k|\eta(x) - \eta(y)|$ for all $x, y \in \mathcal{S}(p)$, then

$$\big\|g_h^{p,\ell,\eta}\big\|_{\mathcal{S}(p),\infty} \le k.$$

In particular, if $h \in L^1(p)$ is Lipschitz continuous with Lipschitz constant 1 then the above applies with $\eta(x) = x$, and $\|g_h^{p,0,\mathrm{Id}}\|_{\mathcal{S}(p),\infty} \le 1$.

3. *(Nonuniform bounds Stein bounds) For all $x \in \mathcal{S}(p)$,*

$$\big|\mathcal{L}_p^\ell f(x)\big| \le 2\|f\|_{\mathcal{S}(p),\infty} \frac{\mathbb{E}[\chi^\ell(X_1, x)]\mathbb{E}[\chi^{-\ell}(x, X_2)]}{p(x)}. \qquad (39)$$

Example 2.34. If $p = \phi$ is the standard Gaussian with cdf Φ, then $\ell = 0$ and the third bound in Proposition 2.33 reduces to $2||f||_\infty \Phi(x)(1 - \Phi(x))/\phi(x)$. The ratio $\Phi(x)(1 - \Phi(x))/\phi(x)$ is closely related to Mill's ratio of the standard normal law. The study of such a function is classical and much is known. For instance, we can apply [10], Theorem 2.3, to get

$$\frac{1}{\sqrt{x^2+4}+x} \le \frac{\Phi(x)(1-\Phi(x))}{\phi(x)} \le \frac{4}{\sqrt{x^2+8}+3x} \tag{40}$$

for all $x \ge 0$. Moreover, $\Phi(x)(1-\Phi(x))/\phi(x) \le \Phi(0)(1-\Phi(0))/\phi(0) = 1/2\sqrt{\pi/2} \approx 0.626$. In particular, Proposition 2.33 recovers the well-known bound $\|\mathcal{L}_p^\ell f\|_\infty \le \sqrt{\pi/2}\|f\|_\infty$; see, for example, [64], Theorem 3.3.1.

3. Covariance identities and inequalities

We start with an easy lower bound inequality, which follows immediately from Lemma 2.4.

Proposition 3.1 (Cramér–Rao type bound). *Let $g \in L^2(p)$. For any $f \in \mathcal{F}_\ell^{(1)}(p)$ such that $\mathcal{T}_p^\ell f \in L^2(p)$ and the assumptions of Lemma 2.4 are satisfied*:

$$\mathrm{Var}\big[g(X)\big] \ge \frac{\mathbb{E}[f(X)(\Delta^{-\ell} g(X))]^2}{\mathbb{E}[(\mathcal{T}_p^\ell f(X))^2]} \tag{41}$$

with equality if and only if there exist α, β real numbers such that $g(x) = \alpha \mathcal{T}_p^\ell f(x) + \beta$ for all $x \in \mathcal{S}(p)$.

Proof. The lower bound (41) follows from the fact that $\mathcal{T}_p^\ell f \in \mathcal{F}^{(0)}(p)$ for all $f \in \mathcal{F}_\ell^{(1)}(p)$ by Lemma 2.3. Therefore, from (14), we have

$$\begin{aligned}\big\{\mathbb{E}\big[f(X)\big(\Delta^{-\ell} g(X)\big)\big]\big\}^2 &= \big\{\mathbb{E}\big[\big(\mathcal{T}_p^\ell f(X)\big) g(X)\big]\big\}^2 \\ &= \big\{\mathbb{E}\big[\big(\mathcal{T}_p^\ell f(X)\big)\big(g(X) - \mathbb{E}\big[g(X)\big]\big)\big]\big\}^2 \\ &\le \mathbb{E}\big[\big(\mathcal{T}_p^\ell f(X)\big)^2\big]\,\mathrm{Var}\big[g(X)\big]\end{aligned}$$

by the Cauchy–Schwarz inequality. □

To obtain upper bounds, we start with an easy consequence of our framework.

Corollary 3.2 (First-order covariance identities). *For all f, g that jointly satisfy the assumptions of Lemma 2.18, we have*

$$\mathrm{Cov}\big[f(X), g(X)\big] = \mathbb{E}\left[\Delta^{-\ell} f(X) \frac{K_p^\ell(X, X')}{p(X)p(X')} \Delta^{-\ell} g\big(X'\big)\right]. \tag{42}$$

Moreover, if choice $f = \mathrm{Id}$ is allowed, then

$$\mathrm{Cov}\big[X, g(X)\big] = \mathbb{E}\big[\tau_p^\ell(X) \Delta^{-\ell} g(X)\big]. \tag{43}$$

Remark 3.3. Identity (42) is provided in [62] (see their equation (11)) in the case $\ell = 0$ for a log-concave density. Some of the history of this identity, including the connection with a classical identity of Hoeffding [50], is provided in [72], Section 2. The earliest version of the same identity (still for $\ell = 0$) we have found in [34], along with applications to measures of correlation as well as further references. A similar identity is provided in [62], without explicit conditions; a clear statement is given in [72], Corollary 2.2, where the identity is proved for absolutely continuous $f \in L^r$ and $g \in L^s$ with conjugate exponents. Our approach shows that it suffices to impose regularity on one of the functions for the identity to hold.

Proof. Let $\bar{f}(x) = f(x) - \mathbb{E}[f(X)]$. Note that $\Delta^\ell \bar{f} = \Delta^\ell f$. To obtain (42), we start from (21) and note that if f, g satisfy the assumptions of Lemma 2.18, then

$$\mathrm{Cov}\big[f(X), g(X)\big] = \mathbb{E}\big[-\big\{\mathcal{L}_p^\ell f(X)\big\}\Delta^{-\ell} g(X)\big]. \tag{44}$$

From this equation, (43) follows immediately. Applying (30), we obtain

$$\mathrm{Cov}\big[f(X), g(X)\big] = \mathbb{E}\left[\mathbb{E}\left[\frac{K_p(X', X)}{p(X')p(X)}\Delta^{-\ell} f\big(X'\big)\Big| X\right]\Delta^{-\ell} g(X)\right]$$

which gives the claim after removing the conditioning. □

Example 3.4. Example 2.16 and identity (43) give the following covariance identities:

- Binomial distribution: For all functions $g : \mathbb{Z} \to \mathbb{R}$ that are bounded on $[0, n]$,

$$\mathrm{Cov}\big[X, g(X)\big] = \mathbb{E}\big[(1-\theta)X\Delta^- g(X)\big] = \theta\mathbb{E}\big[(n-X)\Delta^+ g(X)\big].$$

  Combining the two identities, we also arrive at

$$\mathrm{Cov}\big[X, g(X)\big] = \mathrm{Var}[X]\mathbb{E}\big[\nabla_{\mathrm{bin}(n,\theta)} g(X)\big]$$

  with $\nabla_{\mathrm{bin}(n,\theta)} g(x) = x/n\Delta^- g(x) + (1 - x/n)\Delta^+ g(x)$ the gradient from [48].
- Beta distribution: For all absolutely continuous g such that $\mathbb{E}[|X(1-X)g'(X)|] < \infty$,

$$\mathrm{Cov}\big[X, g(X)\big] = \frac{1}{\alpha+\beta}\mathbb{E}\big[X(1-X)g'(X)\big].$$

It is of interest to work as in [52] to obtain a corresponding upper bound, which would provide some "weighted Poincaré inequality" such as those described in [71]. The representation formulae (42) turns out to simplify the work considerably.

Theorem 3.5. *Fix $h \in L^1(p)$ a decreasing function. For all f, g which satisfy the assumptions of Lemma 2.18, we have*

$$\big|\mathrm{Cov}\big[f(X), g(X)\big]\big| \le \sqrt{\mathbb{E}\left[\big(\Delta^{-\ell} f(X)\big)^2 \frac{-\mathcal{L}_p^\ell h(X)}{\Delta^{-\ell} h(X)}\right]}\sqrt{\mathbb{E}\left[\big(\Delta^{-\ell} g(X)\big)^2 \frac{-\mathcal{L}_p^\ell h(X)}{\Delta^{-\ell} h(X)}\right]} \tag{45}$$

with equality if and only if there exist α_i, $i = 1, \dots, 4$ real numbers such that $f(x) = \alpha_1 h(x) + \alpha_2$ and $g(x) = \alpha_3 h(x) + \alpha_4$ for all $x \in \mathcal{S}(p)$.

Proof. We simply apply (42) and the Cauchy–Schwarz inequality to obtain

$$
\begin{aligned}
\big|\mathrm{Cov}\big[f(X), g(X)\big]\big| &= \left|\mathbb{E}\left[\Delta^{-\ell} f(X) \frac{K_p^\ell(X, X')}{p(X)p(X')} \Delta^{-\ell} g\big(X'\big)\right]\right| \\
&= \left|\mathbb{E}\left[\left\{\frac{\Delta^{-\ell} f(X)}{\sqrt{-\Delta^{-\ell} h(X)}} \sqrt{-\frac{K_p^\ell(X, X')}{p(X)p(X')} \Delta^{-\ell} h\big(X'\big)}\right\}\right.\right. \\
&\quad \left.\left. \times \left\{\frac{\Delta^{-\ell} g(X')}{\sqrt{-\Delta^{-\ell} h(X')}} \sqrt{-\frac{K_p^\ell(X, X')}{p(X)p(X')} \Delta^{-\ell} h(X)}\right\}\right]\right| \\
&\le \sqrt{\mathbb{E}\left[\frac{(\Delta^{-\ell} f(X))^2}{\Delta^{-\ell} h(X)} \frac{K_p^\ell(X, X')}{p(X)p(X')} \Delta^{-\ell} h\big(X'\big)\right]} \\
&\quad \times \sqrt{\mathbb{E}\left[\frac{(\Delta^{-\ell} g(X))^2}{\Delta^{-\ell} h(X)} \frac{K_p^\ell(X, X')}{p(X)p(X')} \Delta^{-\ell} h\big(X'\big)\right]};
\end{aligned}
$$

using (30) leads to the inequality.

The only part of the claim that remains to be proved concerns the saturation condition in the inequality. This follows from the Cauchy–Schwarz inequality, which is an equality if and only if $\Delta^{-\ell} f(x)/\Delta^{-\ell} h(x) \propto \Delta^{-\ell} g(x')/\Delta^{-\ell} h(x')$ is constant throughout $\mathcal{S}(p)$. This is only possible under the stated condition. □

Combining Proposition 3.1 and Theorem 3.5 (applied with $f = g$), we arrive at the following result (applied to a smaller class of functions h) which, as we shall argue below, share a similar flavor to the upper and lower bounds from Theorem 1.1.

Corollary 3.6 (Klaassen bounds, revisited). *For any decreasing function $h \in L^2(p)$ and all g such that Lemma 2.18 applies (with $f = g$), we have*

$$
\begin{aligned}
\frac{\mathbb{E}[-\mathcal{L}_p^\ell h(X)(\Delta^{-\ell} g(X))]^2}{\mathrm{Var}(h(X))} &\le \mathrm{Var}\big[g(X)\big] \\
&\le \mathbb{E}\left[\big(\Delta^{-\ell} g(X)\big)^2 \frac{-\mathcal{L}_p^\ell h(X)}{\Delta^{-\ell} h(X)}\right].
\end{aligned} \tag{46}
$$

Equality in the upper bound holds if and only if there exists constants α, β such that $g(x) = \alpha h(x) + \beta$.

Proof. For the lower bound, we apply Proposition 3.1 with $f(x) = -\mathcal{L}_p^\ell h(x)$ so that $\mathcal{T}_p^\ell c(x) = h(x) - -\mathbb{E}[h(X)]$. For the upper bound, we use Theorem 3.5 with $f = g$. □

Example 3.7 (Pearson and Ord families). Tables 1, 2 and 3 in the supplementary material [38] present the results for distributions which belong to the Pearson and Ord families of distributions. A random variable $X \sim p$ belongs to the *integrated Pearson family* if X is absolutely continuous and there exist $\delta, \beta, \gamma \in \mathbb{R}$ not all equal to 0 such that $\tau_p^\ell(x)(:= -\mathcal{L}_p^\ell(\mathrm{Id})) = \delta x^2 + \beta x + \gamma$ for all $x \in \mathcal{S}(p)$. Similarly, $X \sim p$ belongs to the cumulative Ord family if X is discrete and there exist $\delta, \beta, \gamma \in \mathbb{R}$ not all equal to 0 such that $\tau_p^\ell(x)(:= -\mathcal{L}_p^\ell(\mathrm{Id})) = \delta x^2 + \beta x + \gamma$ for all $x \in \mathcal{S}(p)$. The bounds for these distributions generalize the results, for example, from [3]. For Integrated Pearson distributions, higher order bounds (i.e., bounds in which higher order derivatives of the test functions are considered) are given in [1]. For the cumulative Ord family, we refer to [2] for a detailed study of the associated system of orthogonal polynomials.

Remark 3.8 (About the connection with Klaassen's bounds). The bounds to clarify the connection between the bounds in Corollary 3.6 and those from Theorem 1.1, we follow [52] and restrict our attention to kernels of the form

$$\rho_\zeta^+(x,y) = \mathbb{I}[\zeta < y \le x] - \mathbb{I}[x < y \le \zeta] \quad \text{and}$$
$$\rho_\zeta^-(x,y) = \mathbb{I}[\zeta \le y < x] - \mathbb{I}[x \le y < \zeta]$$

for some $\zeta \in \mathbb{R}$. In our notation, for $\ell \in \{-1, 0, 1\}$,

$$\rho_\zeta^\ell(x,y) = \chi^\ell(\zeta,y)\chi^{-\ell}(y,x) - \chi^\ell(x,y)\chi^{-\ell}(y,\zeta).$$

We first tackle the relation between the main arguments of the bounds, namely $G(x)$ and $g(x)$. Given a measurable function g, we mimic the statement of Theorem 1.1 and introduce the generalized primitive $G(x) = G_\zeta^\ell(x) := \int \rho_\zeta^\ell(x,y)g(y)\mu(\mathrm{d}y) + c$ with c arbitrary, fixed w.l.o.g. to 0. Again in our notation, this becomes

$$G_\zeta^\ell(x) = \int_{\zeta + a_\ell}^{x - b_\ell} g(y)\mu(\mathrm{d}y)\mathbb{I}[\zeta < x] - \int_{x + a_\ell}^{\zeta - b_\ell} g(y)\mu(\mathrm{d}y)\mathbb{I}[x < \zeta].$$

By construction, $\Delta^{-\ell}G_\zeta^\ell(x) = g(x)$ for all ζ and all ℓ, as expected. Nevertheless, in order for $G_\zeta^\ell(x)$ to be well-defined, strong (joint) assumptions on g and ζ are required; for instance, if $g(x) = 1$ then ζ must be finite and $G_{\zeta,c}^\ell(x) = x - \zeta$ while if $g(x)$ has p-mean 0 then the values $\zeta = \pm\infty$ are allowed.

To connect the lower bound (7) and the lower bound of (46), let $k \in L^2(p)$ have p-mean 0. Then

$$\begin{aligned}
\mathbb{E}\big[k(X)\rho_\zeta^\ell(X,x)\big] &= \mathbb{E}\big[k(X)\chi^{-\ell}(x,X)\big]\chi^\ell(\zeta,x) - \mathbb{E}\big[k(X)\chi^\ell(X,x)\big]\chi^{-\ell}(x,\zeta) \\
&= \mathbb{E}\big[k(X)\big]\chi^\ell(\zeta,x) - \mathbb{E}\big[k(X)\chi^\ell(X,x)\big]\big(\chi^\ell(\zeta,x) + \chi^{-\ell}(x,\zeta)\big) \\
&= -\mathbb{E}\big[k(X)\chi^\ell(X,x)\big]
\end{aligned}$$

so that $K(x) = p(x)^{-1}\int \rho_\zeta^\ell(z,x)k(z)p(z)\mu(dz) = -\mathcal{L}_p^\ell k(x)$, and thus (7) follows from the lower bound of (46).

Finally, we consider the upper bounds (6) and (46). Let $H(x) = H^\ell_\zeta(x)$ be a generalized primitive of some nonnegative function h. The same manipulations as above lead to

$$\frac{1}{p(x)}\int \rho^\ell_\zeta(z,x)H(z)p(z)\mu(\mathrm{d}z) = -\mathcal{L}^\ell_p h(x) - \frac{\mathbb{E}[H(X)]}{p(x)}\big(P(x-a_\ell) - \chi^\ell(\zeta,x)\big).$$

If, following [52], we choose ζ in such a way that $\mathbb{E}[H(X)] = 0$ (this is equivalent to requiring $\int_{\zeta+a_\ell}^b h(y)(1-P(y+b_\ell))\mu(\mathrm{d}y) = \int_a^{\zeta-b_\ell} h(y)(P(y-a_\ell))\mu(\mathrm{d}y)$) then we see that the upper bound in (46) is equivalent to (6).

Of course, there is some gain in generality at allowing for a general kernel ρ as in Theorem 1.1, though this comes at the expense of readability: given a positive function h, understanding the form of function H is actually nontrivial and our result illuminates Klaassen's discovery by providing the connection with Stein characterizations.

4. About the weights

The freedom of choice in the test functions h appearing in the bounds invite a study of the impact of the choice of h on the validity and quality of the resulting inequalities.

4.1. Score function and a Brascamp–Lieb inequality

The form of the lower bound in Proposition 3.1 encourages the choice $f(x) = 1$. This is only permitted if the constant function $1 \in \mathcal{F}^{(1)}_\ell(p)$ and $\mathbb{E}[(\mathcal{T}^\ell_p 1(X))^2] < \infty$; these are two strong assumptions which exclude some natural targets such as, for example, the exponential or beta distributions. If this choice is permitted, then we reap the lower bound $\mathrm{Var}[g(X)] \geq (I^\ell(p))^{-1}\mathbb{E}[\Delta^{-\ell}g(X)]$ with $I^\ell(p) = [(\mathcal{T}^\ell_p 1(X))^2]$.

The function $\mathcal{T}^\ell_p 1(x) = \Delta^\ell(p(x))/p(x)$ is some form of generalized score function and $I^\ell(p)$ a generalized Fisher information. Indeed, if $\ell = 0$ and $X \sim p$ is absolutely continuous, then $\mathcal{T}^\ell_p 1(x) = (\log p(x))'$ is exactly the (location) score function of p and $I^{(0)}(p)$ is none other than the (location) Fisher information of p. More generally, we note that if $1 \in \mathcal{F}^{(1)}_\ell(p)$ then $\mathbb{E}[\mathcal{T}^\ell_p 1(X)] = 0$ and, by Lemma 2.4, it satisfies

$$\mathbb{E}\big[\mathcal{T}^\ell_p 1(X)g(X)\big] = -\mathbb{E}\big[\Delta^{-\ell}g(X)\big]$$

for all appropriate g; this further reinforces the analogy.

The corresponding upper bound from (45) is obtained for $h(x) = \mathcal{T}^\ell_p 1(x)$ in (46). Suppose that $p(b^- + a_\ell) = p(a^+ - b_\ell) = 0$. By construction, $\mathcal{L}^\ell_p h(x) = \mathbb{I}_{\mathcal{S}(p)}(x)$. If we can further suppose that $\mathcal{T}^\ell_p 1(x)$ is a decreasing function then

$$\big|\mathrm{Cov}\big[f(X),g(X)\big]\big| \leq \sqrt{\mathbb{E}\bigg[\frac{(\Delta^{-\ell}f(X))^2}{-\Delta^{-\ell}\mathcal{T}^\ell_p 1(X)}\bigg]}\sqrt{\mathbb{E}\bigg[\frac{(\Delta^{-\ell}g(X))^2}{-\Delta^{-\ell}\mathcal{T}^\ell_p 1(X)}\bigg]}.$$

Taking $g = f$, we deduce the following result.

Corollary 4.1 (Brascamp–Lieb inequality). *Under the same conditions as Proposition 2.23, we have*

$$\frac{\mathbb{E}[(\Delta^{-\ell}g(X))]^2}{\mathbb{E}[(\mathcal{T}_p^{\ell}1(X))^2]} \le \mathrm{Var}\big[g(X)\big] \le \mathbb{E}\left[\frac{(\Delta^{-\ell}g(X))^2}{-\Delta^{-\ell}\mathcal{T}_p^{\ell}1(X)}\right] \tag{47}$$

for all g such that $\mathcal{T}_p^{\ell}1$, g satisfy together the assumptions of Lemma 2.18.

Remark 4.2. We refer to (47) as a "Brascamp–Lieb inequality" because of a result from [15] where it is shown that, for a given density p proportional to e^{-V} with V strictly convex on $\mathbb{R}$ and $V' \in L^2(p)$, we have $\mathrm{Var}[g(X)] \le \mathbb{E}[(g'(X))^2/V''(X)]$ where the constant 1 is optimal. Taking $\ell = 0$ in (47), one recognizes $-\Delta^{-\ell}\mathcal{T}_p^{\ell}1 = V''$ so that the upper bound in (47) reduces to this Brascamp–Lieb inequality.

We conclude with a generalized version of the elegant inequality due to [62], Lemma 2.11, in the form stated in [23], equation (1.5).

Corollary 4.3 (Asymmetric Brascamp–Lieb inequality). *Under the same conditions as above, if $-\Delta^{-\ell}\mathcal{T}_p^{\ell}1 \in L^1(\mu)$ then for all f, g in $L^2(p)$,*

$$\big|\mathrm{Cov}\big[f(X), g(X)\big]\big| \le \sup_x \left|\frac{\Delta^{-\ell}f(x)}{\Delta^{-\ell}\mathcal{T}_p^{\ell}1(x)}\right| \mathbb{E}\big[\big|\Delta^{-\ell}g(X)\big|\big]. \tag{48}$$

Proof. Under the assumptions, we may apply (42) to get

$$\begin{aligned}
\big|\mathrm{Cov}\big[f(X), g(X)\big]\big| &\le \mathbb{E}\left[\frac{|\Delta^{-\ell}f(X)|}{-\Delta^{-\ell}\mathcal{T}_p^{\ell}1(X)}\big|\Delta^{-\ell}g(X')\big|\big(-\Delta^{-\ell}\mathcal{T}_p^{\ell}1(X)\big)\frac{K_p^{\ell}(X,X')}{p(X)p(X')}\right] \\
&\le \sup_x \left|\frac{\Delta^{-\ell}f(x)}{\Delta^{-\ell}\mathcal{T}_p^{\ell}1(x)}\right| \mathbb{E}\left[\big|\Delta^{-\ell}g(X')\big|\frac{K_p^{\ell}(X,X')}{p(X)p(X')}\big(-\Delta^{-\ell}\mathcal{T}_p^{\ell}1(X)\big)\right] \\
&= \sup_x \left|\frac{\Delta^{-\ell}f(x)}{\Delta^{-\ell}\mathcal{T}_p^{\ell}1(x)}\right| \mathbb{E}\big[\big|\Delta^{-\ell}g(X')\big|\big]
\end{aligned}$$

where the last line follows by conditioning on X' and applying Proposition 2.23. □

4.2. Stein kernel and Cacoullos' bound

It is natural to consider test function $h = -\mathrm{Id}$ in Theorem 3.5. Since $\Delta^{-\ell}h(x) = 1$, we obtain

$$\big|\mathrm{Cov}\big[f(X), g(X)\big]\big| \le \sqrt{\mathbb{E}\big[\tau_p^{\ell}(X)\big(\Delta^{-\ell}f(X)\big)^2\big]}\sqrt{\mathbb{E}\big[\tau_p^{\ell}(X)\big(\Delta^{-\ell}g(X)\big)^2\big]}.$$

In particular, if $g = f$ then $\mathrm{Var}[g(X)] \le \mathbb{E}[\tau_p^{\ell}(X)(\Delta^{-\ell}g(X))^2]$ in which one recognizes the upper bounds from [16] and also, when $\ell = 0$, [71]. The lower bound in (41) is obtained for

$f(x) = \tau_p^\ell(x)$ for which $\mathcal{T}_p^\ell f(x) = x\mathbb{I}_{\mathcal{S}(p)}(x)$, and the overall bound becomes

$$\frac{\mathbb{E}[\tau_p^\ell(X)(\Delta^{-\ell} g(X))]^2}{\mathrm{Var}[X]} \leq \mathrm{Var}\big[g(X)\big] \leq \mathbb{E}\big[\big(\Delta^{-\ell} g(X)\big)^2 \tau_p^\ell(X)\big]. \tag{49}$$

Example 4.4. In our examples, (49) gives the following covariance identities:

- Binomial distribution: Let $X \sim \mathrm{Bin}(n, \theta)$ as in Example 2.12. From Example 2.16, we obtain the upper and lower bounds

$$\frac{(1-\theta)}{n\theta}\mathbb{E}\big[X\Delta^- g(X)\big]^2 \leq \mathrm{Var}\big[g(X)\big] \leq (1-\theta)\mathbb{E}\big[X\big(\Delta^- g(X)\big)^2\big];$$
$$\frac{\theta}{n(1-\theta)}\mathbb{E}\big[(n-X)\Delta^+ g(X)\big]^2 \leq \mathrm{Var}\big[g(X)\big] \leq \theta\mathbb{E}\big[(n-X)\big(\Delta^+ g(X)\big)^2\big].$$

- Beta distribution: From (3.4), for the Beta(α, β)-distribution with variance $\alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$,

$$\frac{(\alpha+\beta+1)}{\alpha\beta}\mathbb{E}\big[X(1-X)g'(X)\big]^2 \leq \mathrm{Var}\big[g(X)\big]$$
$$\leq \frac{1}{\alpha+\beta}\mathbb{E}\big[X(1-X)\big(g'(X)\big)^2\big].$$

The particular case of other Pearson/Ord distributions is detailed in Tables 1, 2 and 3 available in the supplementary material [38]. The tables include the Binomial distribution and the Beta distribution for easy reference. The Stein operators, which are given, are those from Example 2.17. We remark that Cacoullos' inequality was obtained, in [43], through an argument which is some form of dual to that outlined in this section, for continuous distributions. In that article, the inequality is coined a "weighted Chernov bound"; we refer to [77] as well for extensions towards stable densities.

4.3. Discussion

The theory presented in our paper is closely connected to several classical topics from functional analysis. First, as already mentioned in the Introduction, there is a connection with the *spectral gap and Poincaré inequalities*. Let σ^2 be a positive function on $\mathbb{R}$; the σ^2-weighted Poincaré constant (or spectral gap) of a density p is

$$\lambda_{\sigma^2,p} = \inf_{g \in C_0^\infty(\mathbb{R})} \frac{\mathbb{E}[\sigma^2(X)(g'(X))^2]}{\mathrm{Var}[g(X)]}.$$

The case $\sigma^2 = 1$ leads to the classical (unweighted) Poincaré inequality or spectral gap; if there exists a function g_{opt} achieving equality, one says that the inequality is saturated at g_{opt}. It is an easy matter to use our notation to extend the above to nonabsolutely continuous distributions.

Exploiting the freedom of choice in the test functions h in Corollary 3.6 immediately yields the next result.

Corollary 4.5. *Instate all previous assumptions and notation. Then*

$$\lambda_{\sigma^2,p} \geq \left(\sup_h \sup_x \left\{-\mathcal{L}_p^\ell h(x) / \left(\sigma^2(x) \Delta^{-\ell} h(x)\right\}\right)^{-1}\right.$$

where the supremum is taken over all decreasing functions $h \in L^2(p)$.

It would be of interest to study the connection with the works [13] and [70], and further study the important problem of saturation of the inequalities.

Finally, a connection between variance bounds with Stein's method was already noted, for example, in [35]. It arises naturally in our context by choosing h in Theorem 3.5 such that the corresponding weight $-(\Delta^{-\ell} h(x))^{-1} \mathcal{L}_p^\ell h(x)$ is constant, that is, any mean zero function h such that there exists $\lambda \in \mathbb{R}$ for which

$$\frac{-\mathcal{L}_p^\ell h(x)}{\Delta^{-\ell} h(x)} = \lambda \quad \text{for all } x \in \mathcal{S}(p).$$

By construction and Lemma 2.6, such functions are solution to the eigenfunction problem

$$h(x) = -\lambda \mathcal{T}_p^\ell \left(\Delta^{-\ell} h\right)(x) \quad \text{for all } x \in \mathcal{S}(p)$$

where operator $\mathcal{R}_p^\ell h := \mathcal{T}_p^\ell (\Delta^{-\ell} h)$ is self-adjoint in the sense of that

$$\mathbb{E}\left[\left(\mathcal{R}_p^\ell f(X)\right) g(X)\right] = \mathbb{E}\left[f(X)\left(\mathcal{R}_p^\ell g(X)\right)\right]$$

for all appropriate f, g. We defer a proper treatment to a later publication.

Acknowledgments

The research of Y. Swan was partially supported by the Fonds de la Recherche Scientifique – FNRS under Grant no. MIS F.4539.16. M. Ernst acknowledges partial funding via a Welcome Grant of the Université de Liège and via the Interuniversity Attraction Pole StUDyS (IAP/P7/06). Y. Swan also thanks Lihu Xu for organizing the "Workshop on Stein's method and related topics" at the University of Macau in December 2018, and where a preliminary version of this contribution was first presented. G. Reinert and Y. Swan thank Emilie Clette for fruitful discussions on a preliminary version of this work. Y. Swan thanks Céline Esser for many fruitful discussions. We also thank Benjamin Arras for several pointers to relevant literature, as well as corrections on the first draft of the paper. Finally, we thank the referees and the AE for their remarks and corrections which greatly improved the presentation of this article.

Supplementary Material

Supplement to "First-order covariance inequalities via Stein's method" (DOI: 10.3150/19-BEJ1182SUPP; .pdf). We provide technical proofs from Section 2, several plots of the functions functions $x \mapsto K_p^\ell(x, x')/p(x)$ (for various values of x') and $x \mapsto K_p^\ell(x, x)/p(x)$ and tables for the Pearson and Ord families.

References

[1] Afendras, G. (2013). Unified extension of variance bounds for integrated Pearson family. *Ann. Inst. Statist. Math.* **65** 687–702. MR3094952 https://doi.org/10.1007/s10463-012-0388-3

[2] Afendras, G., Balakrishnan, N. and Papadatos, N. (2018). Orthogonal polynomials in the cumulative Ord family and its application to variance bounds. *Statistics* **52** 364–392. MR3772186 https://doi.org/10.1080/02331888.2017.1406940

[3] Afendras, G., Papadatos, N. and Papathanasiou, V. (2007). The discrete Mohr and Noll inequality with applications to variance bounds. *Sankhyā* **69** 162–189. MR2428867

[4] Afendras, G., Papadatos, N. and Papathanasiou, V. (2011). An extended Stein-type covariance identity for the Pearson family with applications to lower variance bounds. *Bernoulli* **17** 507–529. MR2787602 https://doi.org/10.3150/10-BEJ282

[5] Afendras, G. and Papathanasiou, V. (2014). A note on a variance bound for the multinomial and the negative multinomial distribution. *Naval Res. Logist.* **61** 179–183. MR3197132 https://doi.org/10.1002/nav.21575

[6] Arras, B. and Houdré, C. (2019). *On Stein's Method for Infinitely Divisible Laws with Finite First Moment. SpringerBriefs in Probability and Mathematical Statistics*. Cham: Springer. MR3931309

[7] Arras, B. and Houdré, C. (2019). On Stein's method for multivariate self-decomposable laws. *Electron. J. Probab.* **24** 128. MR4029431

[8] Barbour, A.D., Holst, L. and Janson, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. New York: The Clarendon Press. Oxford Science Publications. MR1163825

[9] Barbour, A.D., Luczak, M.J. and Xia, A. (2018). Multivariate approximation in total variation, II: Discrete normal approximation. *Ann. Probab.* **46** 1405–1440. MR3785591 https://doi.org/10.1214/17-AOP1205

[10] Baricz, Á. (2008). Mills' ratio: Monotonicity patterns and functional inequalities. *J. Math. Anal. Appl.* **340** 1362–1370. MR2390935 https://doi.org/10.1016/j.jmaa.2007.09.063

[11] Bonnefont, M. and Joulin, A. (2014). Intertwining relations for one-dimensional diffusions and application to functional inequalities. *Potential Anal.* **41** 1005–1031. MR3269712 https://doi.org/10.1007/s11118-014-9408-7

[12] Bonnefont, M. and Joulin, A. (2019). A note on eigenvalues estimates for one-dimensional diffusion operators. ArXiv preprint. Available at arXiv:1906.02496.

[13] Bonnefont, M., Joulin, A. and Ma, Y. (2016). A note on spectral gap and weighted Poincaré inequalities for some one-dimensional diffusions. *ESAIM Probab. Stat.* **20** 18–29. MR3519678 https://doi.org/10.1051/ps/2015019

[14] Borovkov, A.A. and Utev, S.A. (1984). On an inequality and a characterization of the normal distribution. *Theor. Probab. Appl.* **28** 219–228.

[15] Brascamp, H.J. and Lieb, E.H. (1976). On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389. MR0450480 https://doi.org/10.1016/0022-1236(76)90004-5

[16] Cacoullos, T. (1982). On upper and lower bounds for the variance of a function of a random variable. *Ann. Probab.* **10** 799–809. MR0659549
[17] Cacoullos, T., Papadatos, N. and Papathanasiou, V. (1997). Variance inequalities for covariance kernels and applications to central limit theorems. *Theor. Probab. Appl.* **42** 1149–155.
[18] Cacoullos, T. and Papathanasiou, V. (1985). On upper bounds for the variance of functions of random variables. *Statist. Probab. Lett.* **3** 175–184. MR0801687 https://doi.org/10.1016/0167-7152(85)90014-8
[19] Cacoullos, T. and Papathanasiou, V. (1986). Bounds for the variance of functions of random variables by orthogonal polynomials and Bhattacharyya bounds. *Statist. Probab. Lett.* **4** 21–23. MR0822720 https://doi.org/10.1016/0167-7152(86)90033-7
[20] Cacoullos, T. and Papathanasiou, V. (1989). Characterizations of distributions by variance bounds. *Statist. Probab. Lett.* **7** 351–356. MR1001133 https://doi.org/10.1016/0167-7152(89)90050-3
[21] Cacoullos, T. and Papathanasiou, V. (1992). Lower variance bounds and a new proof of the central limit theorem. *J. Multivariate Anal.* **43** 173–184. MR1193610 https://doi.org/10.1016/0047-259X(92)90032-B
[22] Cacoullos, T. and Papathanasiou, V. (1995). A generalization of covariance identity and related characterizations. *Math. Methods Statist.* **4** 106–113. MR1324694
[23] Carlen, E.A., Cordero-Erausquin, D. and Lieb, E.H. (2013). Asymmetric covariance estimates of Brascamp–Lieb type and related inequalities for log-concave measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 1–12. MR3060145 https://doi.org/10.1214/11-AIHP462
[24] Chang, W.-Y. and Richards, D.S.P. (1999). Variance inequalities for functions of multivariate random variables. In *Advances in Stochastic Inequalities* (*Atlanta*, *GA*, 1997). *Contemp. Math.* **234** 43–67. Providence, RI: Amer. Math. Soc. MR1694762 https://doi.org/10.1090/conm/234/03444
[25] Chatterjee, S. (2014). A short survey of Stein's method. In *Proceedings of the International Congress of Mathematicians – Seoul* 2014. *Vol. IV* 1–24. Seoul: Kyung Moon Sa. MR3727600
[26] Chatterjee, S. and Shao, Q.-M. (2011). Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie–Weiss model. *Ann. Appl. Probab.* **21** 464–483. MR2807964 https://doi.org/10.1214/10-AAP712
[27] Chen, L.H.Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545. MR0428387 https://doi.org/10.1214/aop/1176996359
[28] Chen, L.H.Y. (1982). An inequality for the multivariate normal distribution. *J. Multivariate Anal.* **12** 306–315. MR0661566 https://doi.org/10.1016/0047-259X(82)90022-7
[29] Chen, L.H.Y. (1985). Poincaré-type inequalities via stochastic integrals. *Z. Wahrsch. Verw. Gebiete* **69** 251–277. MR0779459 https://doi.org/10.1007/BF02450283
[30] Chen, L.H.Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein's Method. Probability and Its Applications* (*New York*). Heidelberg: Springer. MR2732624 https://doi.org/10.1007/978-3-642-15007-4
[31] Chen, P., Nourdin, I. and Xu, L. (2018). Stein's method for asymmetric α-stable distributions, with application to the stable clt. ArXiv preprint. Available at arXiv:1808.02405.
[32] Chernoff, H. (1980). The identification of an element of a large population in the presence of noise. *Ann. Statist.* **8** 1179–1197. MR0594637
[33] Courtade, T.A., Fathi, M. and Pananjady, A. (2019). Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** 777–790. MR3949953 https://doi.org/10.1214/18-aihp898
[34] Cuadras, C.M. (2002). On the covariance between functions. *J. Multivariate Anal.* **81** 19–27. MR1901203 https://doi.org/10.1006/jmva.2001.2000
[35] Diaconis, P. and Zabell, S. (1991). Closed form summation for classical distributions: Variations on a theme of de Moivre. *Statist. Sci.* **6** 284–302. MR1144242

[36] Döbler, C. (2015). Stein's method of exchangeable pairs for the beta distribution and generalizations. *Electron. J. Probab.* **20** 109. MR3418541 https://doi.org/10.1214/EJP.v20-3933
[37] Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11** 7–16. MR1093412 https://doi.org/10.1016/0167-7152(91)90170-V
[38] Ernst, M., Reinert, G. and Swan, Y. (2020). Supplement to "First-order covariance inequalities via Stein's method." https://doi.org/10.3150/19-BEJ1182SUPP.
[39] Ernst, M., Reinert, G. and Swan, Y. (2019). On infinite covariance expansions. ArXiv preprint. Available at arXiv:1906.08376.
[40] Fang, X., Shao, Q.-M. and Xu, L. (2019). Multivariate approximations in Wasserstein distance by Stein's method and Bismut's formula. *Probab. Theory Related Fields* **174** 945–979. MR3980309 https://doi.org/10.1007/s00440-018-0874-5
[41] Fathi, M. (2018). Higher-Order Stein kernels for Gaussian approximation. ArXiv preprint. Available at arXiv:1812.02703.
[42] Fathi, M. (2019). Stein kernels and moment maps. *Ann. Probab.* **47** 2172–2185. MR3980918 https://doi.org/10.1214/18-AOP1305
[43] Furioli, G., Pulvirenti, A., Terraneo, E. and Toscani, G. (2017). Fokker–Planck equations in the modeling of socio-economic phenomena. *Math. Models Methods Appl. Sci.* **27** 115–158. MR3597010 https://doi.org/10.1142/S0218202517400048
[44] Goldstein, L. and Reinert, G. (2005). Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.* **18** 237–260. MR2132278 https://doi.org/10.1007/s10959-004-2602-6
[45] Goldstein, L. and Reinert, G. (2013). Stein's method for the beta distribution and the Pólya–Eggenberger urn. *J. Appl. Probab.* **50** 1187–1205. MR3161381 https://doi.org/10.1239/jap/1389370107
[46] Gorham, J., Duncan, A.B., Vollmer, S.J. and Mackey, L. (2019). Measuring sample quality with diffusions. *Ann. Appl. Probab.* **29** 2884–2928. MR4019878 https://doi.org/10.1214/19-AAP1467
[47] Gorham, J. and Mackey, L. (2017). Measuring sample quality with kernels. In *Proceedings of the* 34*th International Conference on Machine Learning-Volume* 70 1292–1301. JMLR. org.
[48] Hillion, E., Johnson, O. and Yu, Y. (2014). A natural derivative on $[0, n]$ and a binomial Poincaré inequality. *ESAIM Probab. Stat.* **18** 703–712. MR3334010 https://doi.org/10.1051/ps/2014007
[49] Hoeffding, W. (2012). *The Collected Works of Wassily Hoeffding*. New York: Springer.
[50] Höffding, W. (1940). Maszstabinvariante Korrelationstheorie. *Schr. Math. Inst. U. Inst. Angew. Math. Univ. Berlin* **5** 181–233. MR0004426
[51] Karlin, S. (1993). A general class of variance inequalities. In *Multivariate Analysis: Future Directions* (*University Park, PA*, 1992). *North-Holland Ser. Statist. Probab.* **5** 279–294. Amsterdam: North-Holland. MR1246366
[52] Klaassen, C.A.J. (1985). On an inequality of Chernoff. *Ann. Probab.* **13** 966–974. MR0799431
[53] Korwar, R.M. (1991). On characterizations of distributions by mean absolute deviation and variance bounds. *Ann. Inst. Statist. Math.* **43** 287–295. MR1128869 https://doi.org/10.1007/BF00118636
[54] Kusuoka, S. and Tudor, C.A. (2012). Stein's method for invariant measures of diffusions via Malliavin calculus. *Stochastic Process. Appl.* **122** 1627–1651. MR2914766 https://doi.org/10.1016/j.spa.2012.02.005
[55] Landsman, Z., Vanduffel, S. and Yao, J. (2013). A note on Stein's lemma for multivariate elliptical distributions. *J. Statist. Plann. Inference* **143** 2016–2022. MR3095090 https://doi.org/10.1016/j.jspi.2013.06.003
[56] Landsman, Z., Vanduffel, S. and Yao, J. (2015). Some Stein-type inequalities for multivariate elliptical distributions and applications. *Statist. Probab. Lett.* **97** 54–62. MR3299751 https://doi.org/10.1016/j.spl.2014.11.005

[57] Ley, C., Reinert, G. and Swan, Y. (2017). Distances between nested densities and a measure of the impact of the prior in Bayesian statistics. *Ann. Appl. Probab.* **27** 216–241. MR3619787 https://doi.org/10.1214/16-AAP1202

[58] Ley, C., Reinert, G. and Swan, Y. (2017). Stein's method for comparison of univariate distributions. *Probab. Surv.* **14** 1–52. MR3595350 https://doi.org/10.1214/16-PS278

[59] Ley, C. and Swan, Y. (2013). Stein's density approach and information inequalities. *Electron. Commun. Probab.* **18** 7. MR3019670 https://doi.org/10.1214/ECP.v18-2578

[60] Ley, C. and Swan, Y. (2016). Parametric Stein operators and variance bounds. *Braz. J. Probab. Stat.* **30** 171–195. MR3481100 https://doi.org/10.1214/14-BJPS271

[61] Mackey, L. and Gorham, J. (2016). Multivariate Stein factors for a class of strongly log-concave distributions. *Electron. Commun. Probab.* **21** 56. MR3548768 https://doi.org/10.1214/16-ecp15

[62] Menz, G. and Otto, F. (2013). Uniform logarithmic Sobolev inequalities for conservative spin systems with super-quadratic single-site potential. *Ann. Probab.* **41** 2182–2224. MR3098070 https://doi.org/10.1214/11-AOP715

[63] Nash, J. (1958). Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.* **80** 931–954. MR0100158 https://doi.org/10.2307/2372841

[64] Nourdin, I. and Peccati, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein's method to universality. Cambridge Tracts in Mathematics* **192**. Cambridge: Cambridge Univ. Press. MR2962301 https://doi.org/10.1017/CBO9781139084659

[65] Papathanasiou, V. (1995). A characterization of the Pearson system of distributions and the associated orthogonal polynomials. *Ann. Inst. Statist. Math.* **47** 171–176. MR1341214 https://doi.org/10.1007/BF00773421

[66] Prakasa Rao, B.L.S. (2006). Matrix variance inequalities for multivariate distributions. *Stat. Methodol.* **3** 416–430. MR2252395 https://doi.org/10.1016/j.stamet.2005.11.002

[67] Reinert, G. (1995). A weak law of large numbers for empirical measures via Stein's method. *Ann. Probab.* **23** 334–354. MR1330773

[68] Reinert, G., Mijoule, G. and Swan, Y. (2018). Stein gradients and divergences for multivariate continuous distributions. Available at arXiv:1806.03478.

[69] Ross, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132 https://doi.org/10.1214/11-PS182

[70] Roustant, O., Barthe, F. and Iooss, B. (2017). Poincaré inequalities on intervals – application to sensitivity analysis. *Electron. J. Stat.* **11** 3081–3119. MR3694577 https://doi.org/10.1214/17-EJS1310

[71] Saumard, A. (2019). Weighted Poincaré inequalities, concentration inequalities and tail bounds related to Stein kernels in dimension one. *Bernoulli* **25** 3978–4006. MR4010979 https://doi.org/10.3150/19-bej1117

[72] Saumard, A. and Wellner, J.A. (2018). Efron's monotonicity property for measures on $\mathbb{R}^2$. *J. Multivariate Anal.* **166** 212–224. MR3799644 https://doi.org/10.1016/j.jmva.2018.03.005

[73] Saumard, A. and Wellner, J.A. (2019). On the isoperimetric constant, covariance inequalities and L_p-Poincaré inequalities in dimension one. *Bernoulli* **25** 1794–1815. MR3961231 https://doi.org/10.3150/18-BEJ1036

[74] Schoutens, W. (2001). Orthogonal polynomials in Stein's method. *J. Math. Anal. Appl.* **253** 515–531. MR1808151 https://doi.org/10.1006/jmaa.2000.7159

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

[76] Stein, C. (1986). *Approximate Computation of Expectations. Institute of Mathematical Statistics Lecture Notes – Monograph Series* **7**. Hayward, CA: IMS. MR0882007

[77] Toscani, G. (2019). Poincaré-type inequalities for stable densities. *Ric*. *Mat*. **68** 225–236. MR3948329 https://doi.org/10.1007/s11587-018-0398-4
[78] Upadhye, N.S., Čekanavičius, V. and Vellaisamy, P. (2017). On Stein operators for discrete approximations. *Bernoulli* **23** 2828–2859. MR3648047 https://doi.org/10.3150/16-BEJ829
[79] Xu, L. (2019). Approximation of stable law in Wasserstein-1 distance by Stein's method. *Ann*. *Appl*. *Probab*. **29** 458–504. MR3910009 https://doi.org/10.1214/18-AAP1424

Received May 2019 and revised September 2019