Stein-type covariance identities:
Klaassen, Papathanasiou and Olkin-Shepp–type bounds for arbitrary target distributions

Marie Ernst∗, Gesine Reinert† and Yvik Swan∗

December 21, 2018

Abstract

In this paper, following on from [49, 50, 63] we present a minimal formalism for Stein operators which leads to different probabilistic representations of solutions to Stein equations. These in turn provide a wide family of Stein-Covariance identities which we put to use for revisiting the very classical topic of bounding the variance of functionals of random variables. Applying the Cauchy-Schwarz inequality yields first order upper and lower Klaassen [45]-type variance bounds. A probabilistic representation of Lagrange’s identity (i.e. Cauchy-Schwarz with remainder) leads to Papathanasiou [60]-type variance expansions of arbitrary order. A matrix Cauchy-Schwarz inequality leads to Olkin-Shepp [59] type covariance bounds. All results hold for univariate target distribution under very weak assumptions (in particular they hold for continuous and discrete distributions alike). Many concrete illustrations are provided.

1 Introduction

Charles Stein’s mathematical legacy is growing at a remarkable pace and many of the techniques and concepts he pioneered are now a staple of contemporary probability theory. The origins of this stream of research lie in two papers: [67], in which the method was first presented in the context of Gaussian approximation, and [20] where the method was first adapted to a non-Gaussian context, namely that of Poisson approximation. As has been noted by many authors since then, the approach can be applied quasi verbatim to any target distribution other than the Gaussian and the Poisson, under the condition that “correct” ad hoc objects be identified which will permit the basic identities to hold. There now exist several excellent books and reviews on Stein’s method and its consequences in various settings, such as [68, 9, 10, 57, 21]. There also exist several non-equivalent general frameworks for the theory covering to large swaths of probability distributions, of which we single out the works [26, 70] for univariate distributions under analytical assumptions, [6, 7] for infinitely divisible distributions and [55, 35, 37] as well as [29] for multivariate densities under diffusive assumptions. A “canonical” differential Stein operator theory is also presented in [49, 50, 63].

Stein’s method can be broken down into a small number of key steps: [A] identification of a (characterizing) linear operator, [B] bounding of solutions to some differential equations related to this operator, [C] probabilistic Taylor expansions and construction of well-designed couplings; see [62] for an overview. Each of these steps has produced an entire ecosystem of “Stein-type objects” (operators, equations, couplings, etc.). These Stein-type objects are in symbiosis with many classical branches of mathematics such as orthogonal polynomials, functional analysis, PDE theory or Markov chain theory and therefore open bridges between Stein’s theory and these important areas of mathematics. More recently, connections with other more contemporary mathematics have been discovered, such as e.g. information theory as in [58, 8], optimal transportation as in [48, 31], and machine learning as in [36, 53, 23].

In the present paper, we pursue the work begun in [49, 50, 63] and adopt a minimal point of view on all the objects concerned, this time concentrating on the solutions to so-called “Stein equations”.

∗Université de Liège, corresponding author Yvik Swan: yswan@uliege.be.
†University of Oxford.
Aside from its intrinsic interest – which may arguably be only of concern to those meddling directly in the method itself – we shall illustrate the power of our formalism by showing how it allows to obtain optimal and (extremely) flexible upper and lower bounds on arbitrary functionals of random variables with arbitrary univariate distribution. We recover, as particular cases, many (all?) previously known “variance bounds” (also called Poincaré inequalities), including Klaassen’s bounds from [45], Papathanasiou’s variance expansions from [60] (and Houdré and Kagan’s famous expansion from [39]) as well as Olkin and Shepp’s bound from [59]. Our method of proof is in each case new and, moreover, the formalism we introduce makes them in some sense elementary – at least as soon as the framework is laid out. In order to ease the reader into our work, we begin by detailing it in the easiest setting, the formalism we introduce makes them in some sense elementary – at least as soon as the framework is laid out. In order to ease the reader into our work, we begin by detailing it in the easiest setting, namely that of a Gaussian target distribution.

1.1 The Gaussian case

The Gaussian density \( \gamma(x) = (2\pi\sigma^2)^{-1/2}\exp(-(x-\mu)^2/2\sigma^2) \) has many remarkable properties. One of them stems from Stein’s celebrated lemma which reads: a random variable \( N \) has distribution \( N \sim N(\mu, \sigma^2) \) if and only if

\[
\mathbb{E}[(N-\mu)g(N)] = \mathbb{E}[\sigma^2g'(N)]
\]

(1.1)

for all \( g : \mathbb{R} \to \mathbb{R} \) such that \( \mathbb{E}[|g'(N)|] < \infty \). We let \( \mathcal{F}(\gamma) \) be the collection of \( g \) such that \( \mathbb{E}[g'(N)] < \infty \). There are many ways to prove (1.1), but the two basic approaches are:

- **Approach 1**: use the fact that the Gaussian **score function** \( \rho(x) = (\log \gamma(x))' = \gamma'(x)/\gamma(x) = -(x-\mu)/\sigma^2 \) is linear, in combination with integration by parts (see [69]);

- **Approach 2**: use the fact that the Gaussian **Stein kernel** \( \tau(x) = \gamma(x)^{-1}\int_{-\infty}^{x}(u-\mu)\gamma(u)du = \sigma^2 \) is constant, in combination with Fubini’s theorem (see [77] Lemma 1.2]).

Contrarily to appearances, the final result obtained via these two approaches is not identical because (i) there are technical differences concerning the classes of test functions to which the resulting identities apply; (ii) they lead to two formally different identities: where the variance is on the right-hand-side of the equality (1.1) via Approach 1, it is at the denominator of the left hand side of the equality via Approach 2. This is not a simple cosmetic difference that can be brushed away as a byproduct of the standardization, it is rather central to the understanding of the very nature of Stein’s operators and the innocuity of the difference is rather characteristic of the Gaussian distribution. In the language of the present paper, many of the remarkable properties of the Gaussian actually stem from a very “Steinian” characteristic property of the Gaussian: it is the only distribution whose Stein kernel \( \tau(x) = \gamma(x)^{-1}\int_{-\infty}^{x}(u-\mu)\gamma(u)du \) is constant.

In order to exploit Stein’s identity (1.1) in the context of the so-called **Stein’s method**, one starts by considering solutions to the so-called **Stein equations**

\[
\sigma^2f_h'(x) - (x-\mu)f_h(x) = h(x) - \mathbb{E}[h(N)]
\]

(1.2)

where \( h \in \mathcal{H} \) some class of test functions. For any given \( h \in L^1(\gamma) \) there exists (a.s.) a unique solution \( \mathcal{L}_\gamma h = f_h \) to (1.2) such that \( f_h \) is bounded on \( \mathbb{R} \), given by

\[
\mathcal{L}_\gamma h(x) = f_h(x) = \frac{1}{\gamma(x)} \int_{-\infty}^{x} \left( h(u) - \mathbb{E}[h(N)] \right) \gamma(u)du = -\frac{1}{\gamma(x)} \int_{x}^{\infty} \left( h(u) - \mathbb{E}[h(N)] \right) \gamma(u)du.
\]

(1.3)

The operator \( h \mapsto \mathcal{L}_\gamma h \) is called the **pseudo inverse Stein operator**. Such functions as (1.3) can be used in order to assess normality of a real valued random variable \( X \) through the identities:

\[
\sup_{h \in \mathcal{H}} \mathbb{E}[h(X)] - \mathbb{E}[h(N)] = \sup_{h \in \mathcal{H}} \mathbb{E}[\sigma^2f_h'(X) - (X-\mu)f_h(X)] \leq \sup_{f \in \mathcal{F}} \mathbb{E}[\sigma^2f'(X) - (X-\mu)f(X)]
\]

(1.4)

(1.5)

2
where $\mathcal{F} \supset \{ \mathcal{L}_\gamma h \mid h \in \mathcal{H} \}$ is to be “well-chosen”. The expressions on the left hand side of (1.4) are classical, as they correspond to the so-called Integral Probability Metric (IPM): Wasserstein-1 distance for $\mathcal{H}_{\text{Lip}} = \text{Lip}(1)$ the class of all Lipschitz functions with constant 1; Total Variation distance for $\mathcal{H}_{TV} = \{ \mathbb{1}_A, A \subset \mathbb{R} \}$ the class of all indicators of all Borel sets in $\mathbb{R}$; Kolmogorov distance for $\mathcal{H}_{\text{Kol}} = \{ \mathbb{1}_{(-\infty,z]}, z \in \mathbb{R} \}$ the class of indicators of half lines. These are natural measures of probabilistic discrepancy. The right hand side of (1.5) is more mysterious, and one of the secrets of its usefulness lies in the fact that one can chose the class $\mathcal{F}$ to be of a very simple nature. Indeed, we start with the observation that there exist constants $\kappa^i_{\mathcal{H}}, i = 1, 2, \ldots$ such that the functions $f_h$ satisfy the uniform bounds

$$\sup_{h \in \mathcal{H}} \| f_h \|_\infty \leq \kappa^1_{\mathcal{H}}, \quad \sup_{h \in \mathcal{H}} \| f'_h \|_\infty \leq \kappa^2_{\mathcal{H}}, \quad \sup_{h \in \mathcal{H}} \| f''_h \|_\infty \leq \kappa^3_{\mathcal{H}}, \quad \ldots$$

for all important classes $\mathcal{H}$ (including the three mentioned above) – these are the so-called “Stein’s factors”. Given bounds such as (1.6) one can take

$$\mathcal{F} = \{ f : \mathbb{R} \to R \text{ such that } \| f \|_\infty \leq \kappa^1_{\mathcal{H}}, \| f' \|_\infty \leq \kappa^2_{\mathcal{H}}, \text{ etc.} \}$$

so that $\mathcal{F} \supset \{ f_h \text{ of the form } (1.3) \}$ and, crucially, the class $\mathcal{F}$ has a simple structure. This makes (1.5) a very potent starting point for assessing normality of $X$.

Starting from (1.1), it is natural to consider the Stein operator $\mathcal{A}_\gamma f(x) = \sigma^2 f'(x) - (x - \mu) f(x)$ which has the property that $X = N$ (equality in distribution) if and only if $\mathbb{E}[\mathcal{A}_\gamma f(N)] = 0$ for all $f \in \mathcal{F}(\gamma)$. In light of the arguments from the previous paragraph, the inverse of $\mathcal{A}_\gamma$ is given by the pseudo inverse Stein operator which is the integral operator which to any $f \in L^1(\gamma)$ associates the function $\mathcal{L}_\gamma h$ given in (1.3). With this construction, $\mathcal{A}_\gamma \mathcal{L}_\gamma h = h - \mathbb{E}[h[N]]$ for all $h \in L^1(\gamma)$ and $\mathcal{L}_\gamma \mathcal{A}_\gamma h = h$ for all $h \in \mathcal{F}(\gamma)$.

Here are some important examples. With $g_z(x) = \mathbb{1}(x \leq z)$,

$$\mathcal{L}_\gamma g_z(x) = \frac{1}{\gamma(x)} \left( \mathbb{P}(N \leq x \land z) - \mathbb{P}(N \leq x) \mathbb{P}(N \leq z) \right).$$

As a second example, the Stein kernel $\tau(x)$ is nothing other than the solution (1.3) evaluated at $h = \text{Id}$ the identity function; $\mathcal{L}_\gamma \text{Id}(x) = \tau(x)$ which, as already mentioned, is constant and equal to $\sigma^2$ at all $x$. From Stein’s identity we deduce that the Gaussian density is the only density for which the function $\mathcal{L}_\gamma \text{Id}$ is constant. The operator $\mathcal{L}_\gamma$ evaluated at the identity function in general gives the zero bias density from [33], in the sense that $\mathbb{E}[X h(X)] = \mathbb{E}[(\{-\mathcal{L}_\gamma \text{Id}(X)\}) h'(X)] = \mathbb{E}[h'(X^*)]$, with $X^*$ having the $X$-zero bias density.

The Stein operator definition gives in particular that, for $f \in \mathcal{F}(\gamma)$ and $h \in L^1(\gamma)$,

$$\mathbb{E}[(\mathcal{A}_\gamma f(N)) h(N)] = \sigma^2 \mathbb{E}[f'(N) h(N)] - (f h)'(N)] = -\mathbb{E}[f(N) h'(N)].$$

Letting $g(x) = \mathcal{A}_\gamma f(x)$ and using $\mathcal{L}_\gamma \mathcal{A}_\gamma h = h$ allows to generalize (1.1) to the covariance identity:

$$\text{Cov}(g(N), h(N)) = \mathbb{E} \left[ (\mathcal{L}_\gamma g(N)) h'(N) \right]$$

which is valid for all $h, g \in L^1(\gamma)$. For further covariance inequalities we start with the (trivial) observation that

$$\mathcal{L}_\gamma h(x) = \frac{1}{\gamma(x)} \mathbb{E} \left[ (h(N) - \mathbb{E}[h(N)]) \mathbb{1}[N \leq x] \right] = \text{Cov} \left[ h(N), \frac{\mathbb{1}[N \leq x]}{\gamma(x)} \right].$$

Then the following holds.

**Lemma 1.1** (Representation of the inverse Stein operator). Let $N_1, N_2$ be independent copies of $N$. Then

$$-\mathcal{L}_\gamma h(x) = \frac{1}{\gamma(x)} \mathbb{E} \left[ (h(N_2) - h(N_1)) \mathbb{1}[N_1 \leq x \leq N_2] \right]$$

for all $h \in L^1(\gamma)$. 

3
The proof of Lemma 1.1 follows simply by expanding (1.11) and showing that it is equal to (1.10) for all $x$. We will provide details in a (much) more general context in Section 4 (see Lemma 4.1).

Identity (1.11) is not the only available probabilistic representation for $L_\gamma$. The next one we found in [64, Proposition 1].

**Lemma 1.2** (Saumard’s lemma). The symmetric kernel

$$K_\gamma(x, x') = \mathbb{P}[N \leq x \land x'] - \mathbb{P}[N \leq x] \mathbb{P}[N \leq x']$$

is positive definite. Moreover

$$-\mathcal{L}_\gamma h(x) = \mathbb{E}\left[h'(N) \frac{K_\gamma(N, x)}{\gamma(N) \gamma(x)}\right]$$

(1.13)

for all absolutely continuous $h \in L^1(\gamma)$, and letting $N_1, N_2$ be independent copies of $N$ we have

$$\text{Cov}[h(N), g(N)] = \mathbb{E}\left[h'(N_1) \frac{K_\gamma(N_1, N_2)}{\gamma(N_1) \gamma(N_2)} g(N_2)\right]$$

(1.14)

for all absolutely continuous $h, g \in L^2(\gamma)$.

**Remark 1.1.** Functions $x' \mapsto K_\gamma(x, x')/\gamma(x')$ are represented in Figure 1a for various values of $x$.

**Proof.** Symmetry of (1.12) is immediate. To see that it is positive, we note that

$$K_\gamma(x, x') = \begin{cases} \mathbb{P}[N \leq x] \mathbb{P}[N \geq x'] & \text{if } x \leq x' \\ \mathbb{P}[N \leq x'] \mathbb{P}[N \geq x] & \text{if } x > x' \end{cases}$$

To obtain (1.13), we use (1.9) inside (1.10): recalling the notation $g_x(n) = \mathbb{I}[n \leq x]$ we have

$$-\mathcal{L}_\gamma h(x) = \frac{1}{\gamma(x)} \text{Cov}[h(N), g_x(N)] = \frac{1}{\gamma(x)} \mathbb{E}\left[h'(N)\left(-\mathcal{L}_\gamma g_x(N)\right)\right].$$

Using (1.8), with $N'$ an independent copy of $N$,

$$-\mathcal{L}_\gamma g_x(N) = \frac{1}{\gamma(N)} \left(\mathbb{P}(N' \leq x \land N) - \mathbb{P}(N' \leq x) \mathbb{P}(N' \leq N)\right).$$

Hence

$$-\mathcal{L}_\gamma h(x) = \mathbb{E}\left[\frac{1}{\gamma(x) \gamma(N)} \left(\mathbb{P}(N' \leq x \land N) - \mathbb{P}(N' \leq x) \mathbb{P}(N' \leq N)\right) h'(N)\right]$$

and (1.13) follows. For the last point, we simply combine (1.9) and (1.13) again, with $N_2$ yet another independent copy of $N$. Then

$$\text{Cov}[h(N), g(N)] = \mathbb{E}\left[h'(N_1)\left(-\mathcal{L}_\gamma g(N_1)\right)\right] = \mathbb{E}\left[h'(N_2) \mathbb{E}\left[\frac{K_\gamma(N_1, N_2)}{\gamma(N_1) \gamma(N_2)} | N_1\right]\right]$$

and the conclusion follows.

We conclude this introduction by showing how the notations we have introduced are not only of cosmetic value, but that they allow to obtain some powerful results in an efficient manner. For instance, starting from Saumard’s lemma combined with a simple application of the Cauchy-Schwarz inequality, for any test function such that $h' > 0$ a.s. then

$$\text{Var}[g(N)] = \mathbb{E}\left[\frac{g'(N_1) K_\gamma(N_1, N_2)}{\gamma(N_1) \gamma(N_2)} g'(N_2)\right]$$

$$= \mathbb{E}\left[\frac{g'(N_1)^2}{h'(N_1)} \left(\frac{K_\gamma(N_1, N_2) h'(N_2)}{\gamma(N_1) \gamma(N_2)} \frac{K_\gamma(N_1, N_2) h'(N_1)}{\gamma(N_1) \gamma(N_2)} \frac{g'(N_2)}{h'(N_2)}\right)\right]$$

$$\leq \mathbb{E}\left[\frac{g'(N_1)^2}{h'(N_1)} \left(\frac{K_\gamma(N_1, N_2) h'(N_2)}{\gamma(N_1) \gamma(N_2)}\right)^2\right]$$

$$= \mathbb{E}\left[\frac{g'(N_1)^2}{h'(N_1)} \left(\frac{K_\gamma(N_1, N_2) h'(N_2)}{\gamma(N_1) \gamma(N_2)}\right) \left(\frac{K_\gamma(N_1, N_2) h'(N_1)}{\gamma(N_1) \gamma(N_2)}\right)\right]$$

$$= \mathbb{E}\left[\frac{g'(N_1)^2}{h'(N_1)} \left(\frac{K_\gamma(N_1, N_2) h'(N_2)}{\gamma(N_1) \gamma(N_2)}\right) \right] = \mathbb{E}\left[\frac{g'(N_1)^2}{h'(N_1)} \left(-\mathcal{L}_\gamma h(N_1)\right)\right].$$
where we applied (1.13) once again. Lower bounds are just as easy to obtain, from (1.9):

\[
\mathbb{E} \left( \left( -\mathcal{L}_\gamma h(N) \right) g'(N) \right)^2 \leq \mathbb{E} \left( h(N)^2 \right) \text{Var} \left[ g(N) \right]
\]

which leads to the fact that, if \( h \in L^1(\gamma) \) is monotone, then

\[
\frac{\mathbb{E} \left( \left( -\mathcal{L}_\gamma h(N) \right) g'(N)^2 \right)}{\mathbb{E} \left( h(N)^2 \right)} \leq \text{Var} [g(N)] \leq \mathbb{E} \left( \frac{g'(N_1)^2}{h'(N_1)} \right) \left( -\mathcal{L}_\gamma h(N_1) \right).
\]

(1.15)

In particular, taking \( h = \text{Id} \)

\[
\sigma^2 \mathbb{E} \left[ g'(N)^2 \right] \leq \text{Var} [g(N)] \leq \sigma^2 \mathbb{E} \left[ g'(N)^2 \right].
\]

(1.16)

Identity (1.16) is a rephrasing of Chernoff [22]'s classical Gaussian bounds; identity (1.15) is Klaassen’s result [45] in the Gaussian case. As we shall see in Section 4, one can push the argumentation much further and obtain upper and lower infinite variance expansions to any order. In particular we recover the famous variance bound from [39] (which was already available in [60]): for all \( n \geq 1 \),

\[
\sum_{j=1}^{2n}(1)^j \frac{(\sigma^2)^j}{j!} \mathbb{E} \left[ (g^{(j)}(N))^2 \right] \leq \text{Var} [g(N)] \leq \sum_{j=1}^{2n+1}(1)^j \frac{(\sigma^2)^j}{j!} \mathbb{E} \left[ (g^{(j)}(N))^2 \right].
\]

(1.17)

Moreover, we will prove in Section 4 the matrix variance bound:

\[
\begin{pmatrix}
\text{Var} [f(X)] & \text{Cov} [f(X), g(X)] \\
\text{Cov} [f(X), g(X)] & \text{Var} [g(X)]
\end{pmatrix} \leq \sigma^2 \begin{pmatrix}
\mathbb{E} \left[ (f'(X))^2 \right] & \mathbb{E} \left[ f'(X)g'(X) \right] \\
\mathbb{E} \left[ f'(X)g'(X) \right] & \mathbb{E} \left[ (g'(X))^2 \right]
\end{pmatrix}
\]

(inequality in the above indicating that the difference is nonnegative definite), hereby recovering the main result of [59]. Again, our approach applies to basically any univariate target distribution.

### 1.2 Some references

Extensions of Chernoff’s first order bound (1.16) have, of course, attracted much attention. The initiators of the stream of research seem to be [18] and [13], although precursors can be found e.g. in [12]. Chen [18] identified a way to exploit Stein’s operator for the Gaussian distribution not only to simplify Chernoff’s proof, but also to propose a first order upper variance bound for the multivariate Gaussian distribution. [13] identifies the role played by the Stein kernel (and its discrete version) to extend the scope of Chernoff and Chen’s bounds to a very wide class of distributions. A remarkable generalization – and one of the main sources of inspiration behind the current article – is [45] who pinpoints the role plaid by Stein inverse operators in such bounds, and obtains the inequalities in (1.15) for virtually any functional of any univariate probability distribution. Other fundamental early contributions in this topic are [19], [15, 14], or [44] wherein various extensions are proposed (e.g. Karlin deals with the entire class of log-concave distributions). A major breakthrough is due to [60] who obtains infinite expansions for continuous targets, with coefficients very close in spirit to those that we shall propose in Section 4. Papathanasiou’s method of proof in [60] – which rests in an iterative rewriting of the exact remainder in the Cauchy Schwarz identity – is also a direct inspiration for ours. Such results open the way for a successful line of research in connection with Pearson’s and Ord’s system of distributions including works such as [16], [46], [42], [41], and [17]. Similar results, by different means, are Houdré and Kagan’s famous arbitrary order bound (1.17) from [39], extended to wide families of targets e.g. in [41, 40]. The remarkable articles [25] and [47] both propose similar minded considerations in very general settings. More recently, the contributions [4, 5, 6] and [11] and [51, 52] begin to fully explore connections with Stein’s method. In the Gaussian framework, an enlightening first order matrix variance bound is proposed in [59] for the Gaussian distribution, and in a more general setting in [2] (and also to arbitrary order). Finally, we mention [66, 65] and [64]’s revisiting of this classical literature enticed us to begin the work that ultimately led to the present paper. We end this literature review by pointing to [24, 31, 30] wherein fundamental contributions to the theory of Stein kernels are provided in a multivariate setting.
1.3 Outline of the paper and main results

In Section 2 we recall the theory of canonical and standardized Stein operators introduced in [50], and adapt it to our present framework. We introduce a (new) notion of inverse Stein operator (Definition 2.3) along with a first representation formula (2.2); we also introduce one of the basic tools of the paper, namely a generalized indicator function (2.1), which will serve throughout the paper.

In Section 3 we present the main covariance identities (Lemmas 3.1, 3.2) along with a second probabilistic representation of the inverse operator in (3.7). The first variance Klaassen-type bounds are provided in Theorem 3.1. They take on the same exact form as [115]. We stress that the method of proof is new, and elementary. Many standard examples are detailed.

Section 4 begins with a third probabilistic representation of the inverse operator in (4.2), which, despite its simplicity seems to be new. A re-interpretation of the classical Lagrange identity is provided in (4.12) fundamental “iterated Stein coefficients” which we denote $\Gamma_k(x)$; these generalize the classical Stein kernel and we give some abstract examples in Lemma 4.4 with concrete examples also provided for several classical targets. Section 5 details our extension of Olkin and Shepp’s first-order matrix variance-inequality “à la Klaassen”. The final section, Section 6, applies our framework to obtain bounds on solutions of Stein equations.

2 Stein differentiation

Let $X \subset \mathbb{R}$ and equip it with some $\sigma$-algebra $\mathcal{A}$ and $\sigma$-finite measure $\mu$. Let $X$ be a random variable on $\mathcal{X}$, with probability measure $P^X$ which is absolutely continuous with respect to $\mu$; we denote $p$ the corresponding density, and $\mathcal{S}(p) = \{x \in \mathcal{X} : p(x) > 0\}$. Although we could in principle keep the discussion to come very general, in order to make the paper more concrete and readable in the sequel we shall restrict our attention to distributions satisfying the following assumption.

Assumption A. The measure $\mu$ is either the counting measure on $\mathcal{X} = \mathbb{Z}$ or the Lebesgue measure on $\mathcal{X} = \mathbb{R}$. If $\mu$ is the counting measure then there exist $a, b \in \mathbb{Z} \cup \{-\infty, \infty\}$ such that $\mathcal{S}(p) = [a, b] \cap \mathbb{N}$. If $\mu$ is the Lebesgue measure then $\mathcal{S}(p) = [a, b]$ for $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$.

The necessary feature of Assumption A is that $\mu$ is translation-invariant; the assumption of compact support is made for convenience.

Let $\Delta^\ell f(x) = (f(x + \ell) - f(x))/\ell$ for all $\ell \in \mathbb{R}$, with the convention that $\Delta^0 f(x) = f'(x)$, with $f'$ the weak derivative defined Lebesgue almost everywhere. We denote $\text{dom}(p, \Delta^\ell)$ the collection of functions $f : \mathbb{R} \to \mathbb{R}$ such that $\Delta^\ell f(x)$ exists and is finite at all $x \in \mathcal{S}(p)$. The next definitions come from [50].

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

Definition 2.2 (Canonical Stein class). We define $\mathcal{F}^{(0)}(p)$ as the class of functions in $L^1(p)$ which moreover have mean 0 with respect to $p$. The canonical $\ell$-Stein class of $p$ is the collection $\mathcal{F}^{(0)}_{\ell}(p)$ of $f : \mathbb{R} \to \mathbb{R}$ such that $x \mapsto f(x)p(x) \in \text{dom}(p, \Delta^\ell)$ and $T_{\ell}^p f \in \mathcal{F}^{(0)}(p)$ (i.e. $\mathbb{E} [T_{\ell}^p f(X)] = 0$).

Remark 2.1. For details on Stein class and operators, we refer to (i) [50] for the construction in an abstract setting, (ii) [49] for the construction in the continuous setting (i.e. $\ell = 0$) and (iii) [28] for the construction in the discrete setting (i.e. $\ell \in \{-1, 1\}$).
In what follows we restrict attention to the following cases: \( \ell = 0 \) and \( \mu \) is the Lebesgue measure; or \( \ell \in \{-1, +1\} \) and \( \mu \) is counting measure. We note that for positive integer \( \ell \), we have \( \Delta^\ell f(x) = \Delta^1 \left( \frac{1}{\ell} \sum_{k=0}^{\ell-1} f(x + k) \right) \) and with this representation results for other integer values of \( \ell \) could be obtained. For simplicity we do not consider this general case.

By definition, the canonical operators \( T^\ell_p \) embed \( \mathcal{F}^{(1)}_\ell(p) \) into \( \mathcal{F}^{(0)}(p) \) (and in particular \( E[T^\ell_p f(X)] = 0 \) for all \( f \in \mathcal{F}^{(1)}_\ell(p) \)). Inverting this operation is an important part of the construction to come; the following result is well-known and easy to obtain.

**Lemma 2.1** (Representation formula I). Let \( \ell \in \{-1, 0, 1\} \). Let \( p \) with support \( \mathcal{S}(p) \) satisfy Assumption A and define

\[
\chi^\ell(x, y) = \mathbb{I}_{x \leq y - \ell(\ell + 1)/2}.
\]  

(2.1) Then for any \( \eta \in \mathcal{F}^{(0)}(p) \), the function defined on \( \mathcal{S}(p) \) as

\[
\mathcal{L}^\ell_p \eta(x) = \frac{1}{p(x)} \mathbb{E} \left[ \chi^\ell(X, x) \eta(X) \right] = \frac{1}{p(x)} \int \chi^\ell(y, x) \eta(y)p(y)d\mu(y)
\]  

(2.2) satisfies (i) \( \mathcal{L}^\ell_p \eta \in \mathcal{F}^{(1)}_\ell(p) \) and (ii) \( T^\ell_p \mathcal{L}^\ell_p \eta(x) = \eta(x) \) for all \( x \in \mathcal{S}(p) \). If furthermore \( \eta \in \mathcal{F}^{(1)}_\ell(p) \) then also \( \mathcal{L}^\ell_p T^\ell_p \eta(x) = \eta(x) \) for all \( x \in \mathcal{S}(p) \).

**Remark 2.2.** For \( f \in \text{dom}(p, \Delta^\ell) \) such that \( Ef(X) = 0 \) it holds that

\[
\mathcal{L}^\ell_p f(x) = \frac{1}{p(x)} \mathbb{E} \left[ \left( 1 - \chi^\ell(X, x) \right) f(X) \right] = \frac{1}{p(x)} \int \left( 1 - \chi^\ell(y, x) \right) f(y)p(y)d\mu(y).
\]

Hence instead of Definition (2.1) in the developments below a more general version

\[
\zeta^\ell(x, y) = \alpha \mathbb{I}_{x \leq y - \ell(\ell + 1)/2} - \beta \mathbb{I}_{x > y - \ell(\ell + 1)/2}
\]

could be used, where \( \alpha + \beta = 1 \). For simplicity we use Definition (2.1).

The expressions in Lemma 2.1 particularize, in the three cases that interest us, to \( \chi^-(x, y) = \mathbb{I}_{x \leq y} \) \( (\ell = -1) \), \( \chi^+(x, y) = \mathbb{I}_{x < y} \) \( (\ell = 1) \) and \( \chi^0(x, y) = \mathbb{I}_{x \leq y} \) \( (\ell = 0) \). If \( \mu \) is the Lebesgue measure (and \( \ell = 0 \)) then

\[
\mathcal{L}^0_p \eta(x) = \frac{1}{p(x)} \int_a^x \eta(u)p(u)du
\]

(2.3) whereas if \( \mu \) is the counting measure then

\[
\mathcal{L}^+_p \eta(x) = \frac{1}{p(x)} \sum_{j=a}^{x-1} \eta(j)p(j) \quad (\ell = 1) \quad \text{and} \quad \mathcal{L}^-_p \eta(x) = \frac{1}{p(x)} \sum_{j=a}^x \eta(j)p(j) \quad (\ell = -1).
\]

(2.4)

In all cases the functions are extended on \( \mathcal{X} \) through the convention that \( \mathcal{L}^\ell_p \eta(x) = 0 \) for all \( x \notin \mathcal{S}(p) \).

**Definition 2.3** (Canonical pseudo inverse Stein operator). The canonical pseudo-inverse Stein operator is

\[
\mathcal{L}^\ell_p : L^1 \rightarrow \mathcal{F}^{(1)}_\ell(p) : h \mapsto \mathcal{L}^\ell_p (h - Ef(X))
\]

(2.5) with \( \mathcal{L}^\ell_p h(x) \) defined in Lemma 2.1.

Since the operators \( \Delta^\ell \) satisfy the product rule \( \Delta^\ell (f(x)g(x - \ell)) = \Delta^\ell f(x)g(x) + f(x)\Delta^{-\ell}g(x) \), we immediately obtain that \( T^\ell_p (f(x)g(x - \ell)) = T^\ell_p f(x)g(x) + f(x)\Delta^{-\ell}g(x) \) for all appropriate \( f, g \). This leads to
Definition 2.4 (Standardizations of the canonical operator). Fix some $\eta \in \mathcal{F}^{(0)}(p)$. The $\eta$-standardized Stein operator is
\begin{equation}
A_{p}g(x) = T_{p}^{\ell}(L_{p}^{\ell}\eta(\cdot)g(\cdot - \ell))(x) = \eta(x)g(x) + L_{p}^{\ell}\eta(x)(\Delta_{-}\ell g(x))
\end{equation}
acting on the collection $\mathcal{F}(A_{p})$ of test functions $g$ such that $\eta(\cdot)g(\cdot - \ell) \in \mathcal{F}^{(1)}(p)$.

Example 2.1 (Binomial distribution). Stein’s method for the binomial distribution was first developed in [27]. Let $p$ be the binomial density with parameters $(n, p)$. Then $S(p) = [0, n]$, $\mathcal{F}_{+}^{(1)}(p)$ consists of bounded functions such that $f(0) = 0$ and $\mathcal{F}_{-}^{(1)}(p)$ of functions such that $f(n) = 0$. Picking $\ell = 1$ and $\eta(x) = x - np$ then $L_{n,p,\eta}^{\pm}(x) = -(1-p)x$ so that
\begin{equation}
A_{\text{bin}(n,p)}^{1}g(x) = (x-np)g(x) - (1-p)x\Delta^{-}g(x)
\end{equation}
with corresponding class $\mathcal{F}(A_{\text{bin}(n,p)})$ which contains all bounded functions $g : \mathbb{Z} \to \mathbb{R}$. Picking $\ell = -1$ and $\eta(x) = x - np$ then $L_{n,p,\eta}^{\pm}(x) = -p(n-x)$ so that
\begin{equation}
A_{\text{bin}(n,p)}^{2}g(x) = (x-np)g(x) - p(n-x)\Delta^{+}g(x)
\end{equation}
acting on the same class as $[27]$.

Example 2.2 (Poisson distribution). Stein’s method for the Poisson distribution originates in [20]. Let $p$ be the Poisson density with parameter $\lambda$. Then $S(p) = \mathbb{N}$, $\mathcal{F}_{+}^{(1)}(p)$ consists of bounded functions such that $f(0) = 0$ and $\mathcal{F}_{-}^{(1)}(p)$ of functions such that $\lim_{n \to \infty} f(n)\lambda^{n}/n! = 0$. Picking $\ell = 1$ and $\eta(x) = x - \lambda$ then $L_{\lambda,\eta}^{\pm}(x) = -x$ so that
\begin{equation}
A_{\text{Poi}(\lambda)}^{1}g(x) = (x-\lambda)g(x) - x\Delta^{-}g(x)
\end{equation}
acting on $\mathcal{F}(A_{\text{Poi}(\lambda)})$ which contains all bounded functions $g : \mathbb{Z} \to \mathbb{R}$. Picking $\ell = -1$ and $\eta(x) = x-\lambda$ then $L_{\lambda,\eta}^{\pm}(x) = -\lambda$ so that
\begin{equation}
A_{\text{Poi}(\lambda)}^{2}g(x) = (x-\lambda)g(x) - \lambda\Delta^{+}g(x)
\end{equation}
acting on the same class as $[20]$.

Example 2.3 (Beta distribution). Let $p$ be the Beta density with parameters $(\alpha, \beta)$. Then $S(p) = [0, 1]$ and $\mathcal{F}(p)$ consists of functions such that $f(0) = f(1) = 0$. If $\eta(x) = x - \alpha/(\alpha + \beta)$ then $L_{\alpha,\beta,\eta}(x) = -x(1-x)/\alpha + \beta$ leading to the operator
\begin{equation}
A_{\text{Beta}(\alpha,\beta)}g(x) = \left(x - \frac{\alpha}{\alpha + \beta}\right)g(x) - \frac{x(1-x)}{\alpha + \beta}g'(x)
\end{equation}
with corresponding class $\mathcal{F}(A_{\text{Beta}(\alpha,\beta)})$ which contains all differentiable, bounded functions $g : \mathbb{R} \to \mathbb{R}$; see [20] and [37] for more details and related Stein operators.

Example 2.4 (Gamma distribution). Let $p$ be the Gamma density with parameters $(\alpha, \beta)$. Then $S(p) = (0, \infty)$ $(\alpha < 1)$ or $S(p) = [0, \infty)$ $(\alpha \geq 1)$ and $\mathcal{F}(p)$ consists of functions such that $f(0) = 0$ all integrals exist. If $\eta(x) = x - \alpha/\beta$ then $L_{p,\eta}(x) = -\beta x$ leading to the operator
\begin{equation}
A_{\text{Gamma}(\alpha,\beta)}g(x) = (x - \alpha/\beta)g(x) - \beta x g'(x)
\end{equation}
with corresponding class $\mathcal{F}(A_{\text{Gamma}(\alpha,\beta)})$ which contains all differentiable functions $g : \mathbb{R} \to \mathbb{R}$ such that the function $xg(x)$ is bounded; see for example [32] and [54] for more details and related Stein operators.

Example 2.5 (A general example). Let $p$ satisfy Assumption A and suppose that it has finite mean $\mu$. If $\eta(x) = x - \mu$ then $L_{p,\eta}(x) = -\tau_{p}(x)$ is the so-called Stein kernel of $p$, leading to the operator
\begin{equation}
A_{p}g(x) = (x - \mu)g(x) - \tau_{p}(x)g'(x)
\end{equation}
with corresponding class $\mathcal{F}(A_{p})$ which contains all functions $g : \mathbb{R} \to \mathbb{R}$ such that $\tau_{p}g \in \mathcal{F}(p)$. Hence in the general case $[2.6]$ can be viewed as generalisation of the Stein kernel. Implications of this observations are explored in detail in [28].
3 Stein covariance identities and Klaassen-type bounds

The construction from the previous sections is tailored to ensure that we easily deduce the following family of “Stein covariance identities” (e.g. use \( \mathbb{E}[A_{p}^1 g(X)] = 0 \) for \( A_{p}^1 \) as defined in (2.6)). Lemma 3.1 follows directly from the Stein product rule in \([50, \text{Theorem 3.24}]\).

**Lemma 3.1** (Stein IBP formulas). Let \( X \sim p \). Then, for all \( c \in \mathcal{F}_{\ell}^{(1)}(p) \) on the one hand, and all \( \eta \in \mathcal{F}^{(0)}(p) \) on the other hand, we have

\[
\begin{align*}
\mathbb{E} \left[ c(X) \Delta^{-\ell} g(X) \right] &= - \mathbb{E} \left[ \{\mathcal{T}_{p}^\ell c(X)\} g(X) \right] \quad (3.1) \\
\mathbb{E} \left[ - \{\mathcal{L}_{p}^\ell \eta(X)\} \Delta^{-\ell} g(X) \right] &= \mathbb{E} \left[ \eta(X) g(X) \right] \quad (3.2)
\end{align*}
\]

for all \( g \) such that \( c(\cdot)g(\cdot - \ell) \in \mathcal{F}_{\ell}^{(1)}(p) \) (identity (3.1)) and all \( g \) such that \( \mathcal{L}_{p}^\ell \eta(\cdot)g(\cdot - \ell) \in \mathcal{F}_{\ell}^{(1)}(p) \) (identity (3.2)).

These identities provide powerful handles on the target distribution \( X \sim p \). Combining them for instance with a second representation for Stein operator \( \mathcal{L}_{p}^\ell \) given in (3.7) leads to the following result.

**Lemma 3.2** (Representation formula II). Define

\[
K_{p}^\ell(x,x') = \mathbb{E}[\chi^\ell(X,x)\chi^\ell(X,x')] - \mathbb{E}[\chi^\ell(X,x)] \mathbb{E}[\chi^\ell(X,x')].
\]

Then

\[
\begin{align*}
\text{Cov}[h(X),g(X)] &= \mathbb{E} \left[ - \mathcal{L}_{p}^\ell h(X) \Delta^{-\ell} g(X) \right] \quad (3.4) \\
&= \mathbb{E} \left[ \Delta^{-\ell} h(X) \frac{K_{p}^\ell(X,X')}{p(X)p(x')} \Delta^{-\ell} g(x') \right] \quad (3.5) \\
&= \int_{S(p)} \int_{S(p)} \Delta^{-\ell} h(x)K_{p}^\ell(x,x')\Delta^{-\ell} g(x')d\mu(x)d\mu(x') \quad (3.6)
\end{align*}
\]

for all \( h \in L^1(p) \) and \( g \) such that \( c(\cdot)g(\cdot - \ell) \in \mathcal{F}_{\ell}^{(1)}(p) \). Moreover, \( K_{p}^\ell(x,x') \) is positive and bounded. If \( \mathcal{F}^{(0)}(p) \) is dense in \( L^1(p) \), then

\[
- \mathcal{L}_{p}^\ell h(x') = \mathbb{E} \left[ \frac{K_{p}^\ell(X,x')}{p(X)p(x')} \Delta^{-\ell} h(X) \right] \quad (3.7)
\]

for all \( h \in L^1(p) \).

**Proof.** Let \( \tilde{h}(x) = h(x) - \mathbb{E}[h(X)] \). Note that \( \Delta^{\ell} \tilde{h} = \Delta^{\ell} h \). Equality (3.4) follows from (3.2) and the fact that \( \text{Cov}[h(X),g(X)] = \mathbb{E}[(h(X) - \mathbb{E}[h(X)])g(X)] \). To obtain (3.5) we start from (3.4) and use representation (2.2):

\[
\begin{align*}
\text{Cov}[h(X),g(X)] &= \mathbb{E} \left[ - \mathcal{L}_{p}^\ell h(X) \Delta^{-\ell} g(X) \right] = - \mathbb{E} \left[ \tilde{h}(X') \frac{\chi^\ell(X',X)}{p(X)} \Delta^{-\ell} g(X) \right]
\end{align*}
\]

with \( X' \) an independent copy of \( X \). Taking conditional expectations gives

\[
\text{Cov}[h(X),g(X)] = \mathbb{E} \left[ \tilde{h}(X') \frac{\chi^\ell(X',X)}{p(X)} \frac{1}{p(X)} \Delta^{-\ell} g(X) \right].
\]

Next, let \( \rho_x(x') = \chi^\ell(x',x) \) and view this as a function of \( x' \), to obtain from (3.2),

\[
\mathbb{E} \left[ \rho_x(X')\tilde{h}(X') \right] = \mathbb{E} \left[ \tilde{\rho}_x(X')h(X') \right] = \mathbb{E} \left[ - (\mathcal{L}_{p}^\ell \tilde{\rho}_x(X')) \Delta^{-\ell} h(X') \right].
\]
Denoting $\mathcal{L}^\ell_{p,x'}$ as operator $\mathcal{L}^\ell_p$ acting on the first component, indexed by $x'$, we get

$$\text{Cov}[h(X), g(X)] = \mathbb{E}\left[\Delta^{-\ell} h(X') \frac{\mathcal{L}^\ell_{p,x'}(\chi^\ell(X',X))}{p(X)} \Delta^{-\ell} g(X)\right].$$  

(3.8)

Using again (2.2) we have

$$\mathcal{L}^\ell_{p,x'}(\chi^\ell(x',x)) = \frac{1}{p(x')} \mathbb{E}[\chi^\ell(x'',x')(\chi^\ell(x'',x) - \mathbb{E}[\chi^\ell(x'',x)])]$$

$$= \frac{1}{p(x')} \left(\mathbb{E}[\chi^\ell(x'',x')\chi^\ell(x'',x)] - \mathbb{E}[\chi^\ell(x'',x')]\mathbb{E}[\chi^\ell(x'',x)]\right)$$  

(3.9)

with $X''$ another independent copy of $X$. Plugging this last identity into the covariance representation (3.8) we obtain the second equality in (3.5). To see that $K_p^\ell(x,x') \geq 0$ for all $x,x'$ it suffices to notice that $x' \mapsto \chi^\ell(x',x)$ is decreasing and we can apply forthcoming Proposition 6.1. Alternatively, is suffices to note that for all $x,y$,

$$\mathbb{P}(X \leq x \land y) - \mathbb{P}(X \leq x)\mathbb{P}(X \leq y) = \mathbb{P}(X \leq x \land y)(1 - \mathbb{P}(X \leq x \lor y)) \geq 0.$$

Equality 3.7 follows from (3.5) and (2.2) and the assumption that $\mathcal{F}(0)(p)$ is dense in $L^1(p)$. □

Example 3.1 (Binomial distribution). From previous developments we get

$$\text{Cov}[X, g(X)] = \mathbb{E}\left[(1-p)X \Delta^- g(X)\right] = p\mathbb{E}\left[(n-X)\Delta^+ g(X)\right]$$  

(3.10)

which is valid for all functions $g : \mathbb{Z} \to \mathbb{R}$ that are bounded on $[0,n]$. Combining the two identities we also arrive at

$$\text{Cov}[X, g(X)] = \text{Var}[X]\mathbb{E}[\nabla_{\text{bin}(n,p)} g(X)]$$  

(3.11)

with $\nabla_{\text{bin}(n,p)}$ the “natural” binomial gradient $\nabla_{\text{bin}(n,p)} g(x) = \frac{1}{n} \Delta^- g(x) + \frac{n-x}{n} \Delta^+ g(x)$ from [6].

Example 3.2 (Poisson distribution). From previous developments we get

$$\text{Cov}[X, g(X)] = \mathbb{E}\left[X \Delta^- g(X)\right] = \mathbb{E}\left[\lambda \Delta^+ g(X)\right]$$  

(3.12)

for all functions on $\mathbb{Z}$ such that $p(x)g(x) \to 0$ as $x \to \infty$. Similarly as in example 3.1 we combine the two identities to obtain

$$\text{Cov}[X, g(X)] = \text{Var}[X]\mathbb{E}\left[\nabla_{\text{Poi}(\lambda)} g(X)\right],$$  

(3.13)

this time with $\nabla_{\text{Poi}(\lambda)} g(x) = \frac{1}{2} \left(\frac{\lambda}{x} \Delta^- g(x) + \Delta^+ g(x)\right)$.

Example 3.3 (Beta distribution). From previous developments we get

$$\text{Cov}[X, g(X)] = \frac{1}{\alpha + \beta} \mathbb{E}\left[X(1-X)g'(X)\right]$$  

(3.14)

Remark 3.1. Identity 3.9 and the kernel $K_p^0(x,x')$ in (3.3) will turn out very useful in future developments. It is informative to particularize this kernel in three cases of interest to us: letting $P$ denote the cdf of $p$ and $\min(x,x') = x \land x'$, we have

$$K_p^0(x,x') = K_p^-(x,x') = P(x \land x') - P(x)P(x')$$

$$K_p^+(x,x') = P(x \land x' - 1) - P(x - 1)P(x' - 1).$$

Several illustrations are provided in Figure 7.
Theorem 3.1 gives upper and lower bounds of functionals of random variables with coefficients (weights) expressed in terms of Stein operators. This result provides so-called weighted Poincaré inequalities such as those described in [64]; it also contains the main result from [45]. While the lower bound – some type of Cramer-Rao bound – has long been known to have a very simple proof via Stein operators, a proof for the upper bounds has always been either elusive or rather complicated outside of the gaussian case (see e.g. [64]). The representation formulae obtained for Stein integral operators will turn out to simplify the work considerably.

Theorem 3.1 (Klaassen bounds). For any \( c \in \mathcal{F}_\ell (1) \) we have for all \( f \) such that \( \bar{f}^2 \in L^1(p) \):

\[
\text{Var}[f(X)] \geq \frac{E \left[ c(X)(\Delta^{-\ell} f(X)) \right]^2}{E \left[ (T_\ell^p c(X))^2 \right]} \tag{3.15}
\]

with equality if and only if there exist \( \alpha, \beta \) real numbers such that \( f(x) = \alpha T_\ell^p c(x) + \beta \) for all \( x \in S(p) \). Moreover, if \( h \in L^1(p) \) is decreasing then for all \( f \) such that \( \mathcal{L}_\ell^p h(\cdot)f(\cdot - \ell) \in \mathcal{F}_\ell (1) \) we have

\[
\text{Var}[f(X)] \leq E \left[ (\Delta^{-\ell} f(X))^2 - \mathcal{L}_\ell^p h(X) \right] \tag{3.16}
\]

with equality if and only if there exist \( \alpha, \beta \) real numbers such that \( f(x) = \alpha h(x) + \beta \) for all \( x \in S(p) \).

Remark 3.2. The requirement that \( h \) be decreasing is not necessary; all that is needed is in fact that \( -\mathcal{L}_\ell^p h(X)/\Delta^{-\ell} h(X) > 0 \) almost surely.

Proof of Theorem 3.1. The lower bound (3.15) is an almost immediate consequence of (3.1). Indeed, using the fact that \( T_\ell^p c(X) \in \mathcal{F}(0) \) for all \( c \in \mathcal{F}(p) \) we have

\[
E \left[ c(X)(\Delta^{-\ell} f(X))^2 \right] = E \left[ (T_\ell^p c(X)) f(X) \right]^2 = E \left[ (T_\ell^p c(X)) \bar{f}(X) \right]^2 \leq E \left[ (T_\ell^p c(X))^2 \right] \text{Var}[f(X)]
\]
by the Cauchy-Schwarz inequality. To see (3.16) we simply apply (3.5) and the Cauchy-Schwarz inequality to obtain

\[
\text{Var}[f(X)] = E \left[ \tilde{f}(X) f(X) \right] = E \left[ \Delta^{-\ell} f(X) \frac{K^\ell_p(X,X')}{p(X)p(X')} \Delta^{-\ell} f(X') \right]
\]

\[
= E \left\{ \Delta^{-\ell} f(X) \sqrt{\Delta^{-\ell} h(X)} \frac{K^\ell_p(X,X')}{p(X)p(X')} \Delta^{-\ell} f(X') \right\} \left\{ \Delta^{-\ell} f(X') \sqrt{\Delta^{-\ell} h(X')} \frac{K^\ell_p(X,X')}{p(X)p(X')} \Delta^{-\ell} h(X) \right\}
\]

\[
\leq E \left[ \frac{(\Delta^{-\ell} f(X))^2 K^\ell_p(X,X')}{\Delta^{-\ell} h(X)} \Delta^{-\ell} h(X') \right] ;
\]

using (3.7) finishes the proof. \( \Box \)

Example 3.4 (Binomial bounds). Let \( X \sim \text{Bin}(n,p) \). From previous developments we obtain the upper and lower bounds

\[
\frac{n}{np} \text{Var}[X] ^{X} \leq \text{Var}[f(X)] \leq (1-p) \text{Var}[X] ^{X} (\Delta^{-\ell} f(X))^2 ;
\]

\[
\frac{n}{n(1-p)} \text{Var}[X] ^{X} \leq \text{Var}[f(X)] \leq \text{Var}[p(n-X)(\Delta^{-\ell} f(X))^2] .
\]

Example 3.5 (Poisson bounds). Let \( X \sim \text{Pois}(\lambda) \). From previous developments we obtain the upper and lower bounds

\[
\frac{\lambda}{\text{Var}[f(X)]} \leq \text{Var}[f(X)] \leq \text{Var}[\text{Var}[f(X)](\Delta^{-\ell} f(X))^2] .
\]

Example 3.6 (Beta bounds). Let \( X \sim \text{Beta}(\alpha, \beta) \). From previous developments we obtain the upper and lower bounds

\[
\frac{(\alpha + \beta)(\alpha + \beta + 1)}{\alpha \beta} \text{Var}[X(1-X)f'(X)] \leq \text{Var}[f(X)] \leq \frac{1}{\alpha + \beta} \text{Var}[X(1-X)(f'(X))^2] .
\]

4 A Lagrange formula and Papathanassiou-type variance expansions

This section begins with a third representation for the Stein operator \( L^\ell_p \); surprisingly, this result seems to be new.

Lemma 4.1 (Representation formula III). Let \( X, X_1, X_2 \) be independent copies of \( X \sim p \) with support \( S(p) \). We define

\[
\Phi^\ell_p(u, v) = \chi^\ell(u, x)\chi^{-\ell}(x, v)/p(x)
\]

for all \( u, v \in S(p) \) (note that \( \Phi^\ell_p(u, x, v) = 0 \) for \( u \geq v \)). Then, for all \( h \in L^1(p) \) we have

\[
- L^\ell_p h(x) = E \left[ (h(X_2) - h(X_1))\Phi^\ell_p(X_1, x, X_2) \right] .
\]

Proof. First notice that, under the assumptions, \( E[\chi^\ell(X, x) + \chi^{-\ell}(x, X)] = 1 \) for all \( x \in X \). Suppose without loss of generality that \( E[h(X)] = 0 \). Using that \( X_1, X_2 \) are i.i.d., we reapply

\[
E[h(X_1) \chi^\ell(X_1, x)] = E[h(X_2) \chi^\ell(X_2, x)]
\]

\[
= E[h(X_2) \chi^\ell(X_2, x)] - E[h(X_2)\chi^{-\ell}(x, X_2)] E[\chi^\ell(X_1, x)]
\]

\[
= E[h(X_1) \chi^\ell(X_1, x)] E[\chi^{-\ell}(x, X_2)] - E[h(X_2)(1 - \chi^\ell(X_2, x)))] E[\chi^\ell(X_1, x)]
\]

\[
= E[h(X_1) \chi^\ell(X, x)] (E[\chi^{-\ell}(x, X)] + E[\chi^\ell(X, x)]) .
\]

The conclusion follows after recalling (2.2). \( \Box \)
Combining (3.4) and (1.2) we easily obtain the following identities which, together, will ultimately lead to our infinite expansions for variances.

**Lemma 4.2.** Let $X \sim p$ with support $S(p)$. For all $x_1, x_2 \in S(p)$, we have

$$g(x_2) - g(x_1) = \mathbb{E} \left[ \Phi_p^\ell(x_1, X, x_2) \Delta^{-\ell} g(X) \right].$$

(4.3)

If $X_1, X_2$ are independent copies of $X$ then

$$\text{Var}[g(X)] = \mathbb{E} \left[ (g(X_2) - g(X_1))^2 I_{[X_1 < X_2]} \right]$$

(4.4)

for all $g \in L^2(p)$.

**Proof.** Identity (4.3) follows from the (trivial) observation that

$$\mathbb{E} \left[ \frac{\chi^\ell(x_1, X) \chi^{-\ell}(X, x_2)}{p(X)} \Delta^{-\ell} g(X) \right] = \int_{S(p)} \chi^\ell(x_1, x) \chi^{-\ell}(x, x_2) \Delta^{-\ell} g(x) d\mu(x)$$

which is equal to $g(x_2) - g(x_1)$ for all $x_1 < x_2 \in S(p)$ and all $\ell \in \{-1, 0, 1\}$. Using (3.4) followed by (4.2), we get

$$\text{Var}[g(X)] = \mathbb{E}[-\mathcal{I}^\ell_p g(X) \Delta^{-\ell} g(X)]$$

$$= \mathbb{E} \left[ (g(X_2) - g(X_1)) \Phi_p^\ell(X_1, X, X_2) | X \right] \Delta^{-\ell} g(X)$$

$$= \mathbb{E} \left[ (g(X_2) - g(X_1)) \Phi_p^\ell(X_1, X, X_2) \Delta^{-\ell} g(X) \right]$$

$$= \mathbb{E} \left[ (g(X_2) - g(X_1)) \mathbb{E} \left[ \Phi_p^\ell(X_1, X, X_2) \Delta^{-\ell} g(X) | X_1, X_2 \right] I_{[X_1 < X_2]} \right]$$

$$= \mathbb{E} \left[ (g(X_2) - g(X_1))^2 I_{[X_1 < X_2]} \right],$$

whence the claim. \hfill \Box

**Remark 4.1.** In fact, treating the discrete and continuous cases separately, one could also obtain identity (4.4) as a direct application of Lagrange’s identity (a.k.a. the Cauchy-Schwarz inequality with remainder) which reads, in the finite discrete case, as

$$\left( \sum_{k=u}^v a_k^2 \right) \left( \sum_{k=u}^v b_k^2 \right) - \left( \sum_{k=u}^v a_k b_k \right)^2 \leq \sum_{i=u}^{v-1} \sum_{j=i+1}^v (a_i b_j - a_j b_i)^2$$

(4.5)

with $a_k = g(k)\sqrt{p(k)}$ and $b_k = \sqrt{p(k)}$ for $k = 0, \ldots, n$. Identity (4.5) and its continuous counterpart will play a crucial role in the sequel; it will be more suited to our cause under the following form.

**Lemma 4.3** (A probabilistic Lagrange identity). Let $X, X_1, X_2$ be independent random variables with distribution $p$ and $a, b$ any two functions such that the expectations below exist. Then

$$\mathbb{E} \left[ a(X) b(X) \Phi_p^\ell(u, X, v) \right]^2 = \mathbb{E} \left[ (a(X))^2 \Phi_p^\ell(u, X, v) \right] \mathbb{E} \left[ (b(X))^2 \Phi_p^\ell(u, X, v) \right]$$

$$- \mathbb{E} \left[ (a(X)_1 b(X)_2 - a(X)_2 b(X)_1))^2 \Phi_p^\ell(u, X_1, X_2, v) \right]$$

(4.6)

where

$$\Phi_p^\ell(u, x_1, x_2, v) = \frac{\chi^\ell(u, x_1) \chi^\ell(x_1, x_2) \chi^{-\ell}(x_2, v)}{p(x_1)p(x_2)}.$$

(4.7)
Proof of Lemma 4.3. The proof follows from expanding the second term in (4.6),

\[ \mathbb{E} \left[ (a(X_1)b(X_2) - a(X_2)b(X_1))^2 \Phi_p^\ell(u, X_1, X_2, v) \right] \]

\[ = \mathbb{E} \left[ (a(X_1)b(X_2))^2 + (a(X_2)b(X_1))^2 \Phi_p^\ell(u, X_1, v)\Phi_p^\ell(u, X_2, v) \Delta^\ell(X_1, X_2) \right] \]

\[-2\mathbb{E} \left[ (a(X_1)b(X_2)a(X_2)b(X_1))\Phi_p^\ell(u, X_1, v)\Phi_p^\ell(u, X_2, v) \Delta^\ell(X_1, X_2) \right].\]

By symmetry,

\[ \mathbb{E} \left[ (a(X_1)b(X_2) - a(X_2)b(X_1))^2 \Phi_p^\ell(u, X_1, X_2, v) \right] \]

\[ = \mathbb{E} \left[ (a(X_1)b(X_2) - a(X_2)b(X_1))^2 \Phi_p^\ell(u, X_2, X_1, v) \right].\]

Now \( \Delta^\ell(X_1, X_2) + \Delta^\ell(X_2, X_1) = 1 \) when \( X_1 = X_2 \), and vanishes otherwise, and for \( X_1 = X_2 \), it holds that \( (a(X_1)b(X_2) - a(X_2)b(X_1))^2 = 0 \). Hence

\[ \mathbb{E} \left[ (a(X_1)b(X_2) - a(X_2)b(X_1))^2 \Phi_p^\ell(u, X_1, X_2, v) \right] \]

\[ = \frac{1}{2} \mathbb{E} \left[ (a(X_1)b(X_2) - a(X_2)b(X_1))^2 \Phi_p^\ell(u, X_1, v)\Phi_p^\ell(u, X_2, v) \right].\]

Exploiting the independence of \( X_1 \) and \( X_2 \) now yields the conclusion. \( \square \)

Remark 4.2. For ease of reference, we detail (4.7):

\[ \Phi_p^0(u, x_1, x_2, v) = \Phi_p^0(u, x_1, x_2) \frac{\chi^0(x_2, v)}{p(x_2)} = \frac{\chi^0(u, x_1)\chi^0(x_1, x_2)\chi^0(x_2, v)}{p(x_1)p(x_2)} \] (4.8)

\[ \Phi_p^- (u, x_1, x_2, v) = \Phi_p^- (u, x_1, x_2) \frac{\chi^+ (x_2, v)}{p(x_2)} = \frac{\chi^-(u, x_1)\chi^+ (x_1, x_2)\chi^+ (x_2, v)}{p(x_1)p(x_2)} \] (4.9)

\[ \Phi_p^+ (u, x_1, x_2, v) = \frac{\chi^+ (u, x_1)}{p(x_1)} \Phi_p^+ (x_1, x_2, v) = \frac{\chi^+ (u, x_1)\chi^+ (x_1, x_2)\chi^- (x_2, v)}{p(x_1)p(x_2)} \] (4.10)

Theorem 4.1. Fix \( \ell \in \{-1, 0, 1\} \) and define the sequence \( \ell_n = (\ell_n)_{n \geq 1} \) such that \( \ell_n = 0 \) for all \( n \) if \( \ell = 0 \), otherwise \( \ell_n \in \{-1, 1\} \) arbitrarily chosen. Consider a sequence \( (h_n)_{n \geq 1} \) of real valued functions \( h_i : \mathbb{R} \to \mathbb{R} \) such that \( \mathbb{P}[\Delta^\ell h_{2i-1}(X) > 0] = 1 \) for all \( i \geq 1 \). Starting with some function \( g \), we also recursively define the sequence \( (g_n)_{n \geq 1} \) by \( g_0(x) = g(x) \) and \( g_i(x) = \Delta^\ell g_i(x) / \Delta^\ell h_i(x) \) for all \( x \in S(p) \). Finally, starting from \( \Phi_p^\ell (x_1, x, x_2) = \Phi_p^\ell (x_1, x, x_2) \) and \( \Phi_p^\ell (x_1, x, x_2, x_2) \) we define recursively for \( n \geq 2 \)

\[ \Phi_p^\ell (x_1, x, x_2, \ldots, x_{2n-1}, x, x_{2n}, \ldots, x_2) \]

\[ = \Phi_p^\ell (x_{2n-1}, x, x_{2n}) \Phi_p^\ell (x_{2n-2}, x, x_{2n+1}, \ldots, x_2) \] (4.11)

\[ \Phi_p^\ell (x_1, x, x_2, \ldots, x_{2n-1}, x_{2n+1}, x_{2n+2}, \ldots, x_2) \]

\[ = \Phi_p^\ell (x_{2n-1}, x_{2n+1}, x_{2n+2}) \Phi_p^\ell (x_{2n-2}, x_{2n+1}, \ldots, x_2) \] (4.12)

at any sequence \( (x_j)_{j \geq 1} \). Then, for all \( g \in L^2(p) \) and all \( n \geq 1 \) we have

\[ \text{Var}[g(X)] = \sum_{k=1}^{n} (-1)^{k-1} \mathbb{E} \left[ \frac{(\Delta^{\ell_k} g_{k-1}(X))^2}{\Delta^{\ell_k} h_k(X)} \Gamma_k^\ell(X) \right] + (-1)^n R_n^\ell \] (4.13)

where

\[ \Gamma_k^\ell(x) = \mathbb{E} \left[ (h_k(X_{2k}) - h_k(X_{2k-1})) \prod_{i=1}^{k-1} \Delta^{\ell_i} h_i(X_{2i+1}, X_{2i+2}) \Phi_{k,1}^\ell (X_1, \ldots, X_{2k-1}, x, X_{2k}, \ldots, X_2) \right] \] (4.14)
We treat the two terms separately. First, variable function $\Psi(x) = g_n(x) - g_n(x_{n+1})$ taken only with respect to $X_1, \ldots, X_n$ (that is, conditionally on all non-mentioned variables). Starting from (4.14) and using (4.13):

$$
\begin{align*}
\text{Var}[g(X)] &= E \left[ (g(X_2) - g(X_1))^2 \mathbb{I}_{[X_1 < X_2]} \right] \\
&= E \left[ E_X \left[ \Phi_p(X_1, X_2) \Delta^{-\ell_1} g(X) \right]^2 \mathbb{I}_{[X_1 < X_2]} \right].
\end{align*}
$$

(we introduce the notation $\Delta\ell h_k(x, y) = \Delta\ell h_k(x)\Delta\ell h_k(y)$ and an empty product is set to 1).

**Proof.** Throughout this proof we write $E_{X_1, \ldots, X_n} [\Psi(X_1, \ldots, X_n)]$ to denote the expectation of multivariable function $\Psi(X_1, \ldots, X_n)$ taken only with respect to $X_1, \ldots, X_n$ (that is, conditionally on all non-mentioned variables). Starting from (4.14) and using (4.13):

$$
\begin{align*}
\text{Var}[g(X)] &= E \left[ (g(X_2) - g(X_1))^2 \mathbb{I}_{[X_1 < X_2]} \right] \\
&= E \left[ E_X \left[ \Phi_p(X_1, X, X_2) \Delta^{-\ell_1} g(X) \right]^2 \mathbb{I}_{[X_1 < X_2]} \right].
\end{align*}
$$

Next, for any $h_1$ such that $P[\Delta^{-\ell_1} h_1(X) > 0] = 1$, we have, thanks to (4.16) and conditionally on $X_1, X_2$:

$$
\begin{align*}
\left( E_X \left[ \Phi_p(X_1, X, X_2) \Delta^{-\ell_1} g(X) \right] \right)^2 &= \left( E_X \left[ \Phi_p(X_1, X, X_2) \frac{\Delta^{-\ell_1} g(X)}{\sqrt{\Delta^{-\ell_1} h_1(X)}} \sqrt{\Delta^{-\ell_1} h_1(X)} \right] \right)^2 \\
&= E_X \left[ \frac{\Delta^{-\ell_1} g(X)}{\sqrt{\Delta^{-\ell_1} h_1(X)}} \Phi_p(X_1, X, X_2) \right] E_{X'} \left[ \Delta^{-\ell_1} h_1(X') \Phi_p(X_1, X', X_2) \right] \\
&- E_{X_1, X_2} \left[ \frac{\Delta^{-\ell_1} g(X_3)}{\sqrt{\Delta^{-\ell_1} h_1(X_3)}} \frac{\Delta^{-\ell_1} h_1(X_4)}{\sqrt{\Delta^{-\ell_1} h_1(X_4)}} \frac{\Delta^{-\ell_1} g(X_4)}{\sqrt{\Delta^{-\ell_1} h_1(X_4)}} \Phi_p(X_1, X_3, X_4, X_2) \right] \\
&=: I_1(X_1, X_2) - I_2(X_1, X_2)
\end{align*}
$$

($X'$ in the second line is another independent copy of $X$). We need to compute $E[I_1(X_1, X_2) \mathbb{I}_{[X_1 < X_2]}] - E[I_2(X_1, X_2) \mathbb{I}_{[X_1 < X_2]}]$. We begin by noting that, in the discrete case, the strict inequality in the indicator $\mathbb{I}_{[X_1 < X_2]}$ is implicit in $\Phi_p(X_1, X, X_2) = \chi(X_1, X) \chi^{-\ell_1}(X, X_2) / p(X)$ and hence a fortiori also in $\Phi_p(X_1, X, X_2)$. In the continuous case there is no difference between $\mathbb{I}_{[X_1 < X_2]}$ and $\mathbb{I}_{[X_1 < X_2]}$. We treat the two terms separately. First,

$$
\begin{align*}
E[I_1(X_1, X_2) \mathbb{I}_{[X_1 < X_2]}] &= E \left[ E_X \left[ \frac{\Delta^{-\ell_1} g(X)}{\sqrt{\Delta^{-\ell_1} h_1(X)}} \Phi_p(X_1, X, X_2) \right] E_{X'} \left[ \Delta^{-\ell_1} h_1(X') \Phi_p(X_1, X', X_2) \right] \mathbb{I}_{[X_1 < X_2]} \right] \\
&= E \left[ \frac{\Delta^{-\ell_1} g(X)}{\sqrt{\Delta^{-\ell_1} h_1(X)}} \Phi_p(X_1, X, X_2) \right] E_{X'} \left[ \Delta^{-\ell_1} h_1(X') \Phi_p(X_1, X', X_2) \right] \\
&= E \left[ \frac{\Delta^{-\ell_1} g(X)}{\sqrt{\Delta^{-\ell_1} h_1(X)}} (h_1(X_2) - h_1(X_1)) \Phi_p(X_1, X, X_2) \right],
\end{align*}
$$

the last identity by (4.13). For the second term we have

$$
\begin{align*}
E[I_2(X_1, X_2) \mathbb{I}_{[X_1 < X_2]}] &= E \left[ \frac{\Delta^{-\ell_1} g(X_3)}{\sqrt{\Delta^{-\ell_1} h_1(X_3)}} \frac{\Delta^{-\ell_1} g(X_4)}{\sqrt{\Delta^{-\ell_1} h_1(X_4)}} \Phi_p(X_1, X_3, X_4, X_2) \right] \\
&= E \left[ \Delta^{-\ell_1} h_1(X_3) \Delta^{-\ell_1} h_1(X_4) \left( g_1(X_4) - g_1(X_3) \right)^2 \Phi_p(X_1, X_3, X_4, X_2) \right].
\end{align*}
$$

(4.19)
with \( g_1(x) = \Delta^{-\ell_1} g(x)/\Delta^{-\ell_1} h_1(x) \), as anticipated. Combining (4.18) and (4.19) we obtain (4.13) at \( n = 1 \).

To pursue towards \( n = 2 \), starting from (4.19), we simply apply the same process as above:

\[
\mathbb{E} \left[ \Delta^{-\ell_1} h_1(X_3, X_4) (g_1(X_3) - g_1(X_3))^2 \Phi_{1,2}(X_1, X_3, X_4, X_2) \right] \\
= \mathbb{E} \left[ \Delta^{-\ell_1} h_1(X_3, X_4) \left( \mathbb{E}_X \left[ \Phi_{\ell_2}^2(X_3, X, X_4) \Delta^{-\ell_2} g_1(X) \right] \right)^2 \Phi_{1,2}(X_1, X_3, X_4, X_2) \right] \\
= \mathbb{E} \left[ \Delta^{-\ell_1} h_1(X_3, X_4) I_{21}(X_3, X_4) \Phi_{1,2}(X_1, X_3, X_4, X_2) \right] \\
- \mathbb{E} \left[ \Delta^{-\ell_1} h_1(X_3, X_4) I_{22}(X_3, X_4) \Phi_{1,2}(X_1, X_3, X_4, X_2) \right]
\]

with \( I_{2j}, j = 1, 2 \) defined (in terms of a function \( h_2 \)) through (4.17). Again we treat the terms separately:

\[
\mathbb{E} \left[ \Delta^{-\ell_1} h_1(X_3, X_4) I_{21}(X_3, X_4) \Phi_{1,2}(X_1, X_3, X_4, X_2) \right] \\
= \mathbb{E} \left[ \frac{(\Delta^{-\ell_2} g_1(X))^2}{\Delta^{-\ell_2} h_2(X)} (h_2(X_4) - h_2(X_3)) \Delta^{-\ell_1} h_1(X_3, X_4) \Phi_{\ell_2}^2(X_3, X, X_4) \Phi_{1,2}(X_1, X_3, X_4, X_2) \right]
\]

and

\[
\mathbb{E} \left[ \Delta^{-\ell_1} h_1(X_3, X_4) I_{22}(X_3, X_4) \Phi_{1,2}(X_1, X_3, X_4, X_2) \right] \\
= \mathbb{E} \left[ \Delta^{-\ell_1} h_1(X_3, X_4) \Delta^{-\ell_2} h_2(X_5, X_6) (g_2(X_6) - g_2(X_5))^2 \Phi_{\ell_2}^2(X_3, X_5, X_6, X_4) \Phi_{1,2}(X_1, X_3, X_4, X_2) \right] .
\]

The complete statement follows by induction and careful bookkeeping of all the indices. □

**Remark 4.3** (About the assumptions in the theorem). A stronger sufficient condition on the functions \( h_i \) is that they be strictly increasing throughout \( S(p) \), in which case the condition \( \Delta^i h(x) > 0 \) is guaranteed. The prototypical example of such a sequence is \( h_i(x) = x \) for all \( i \geq 1 \) which clearly satisfies all the required assumptions. When \( \ell \neq 0 \) then the condition that \( \mathbb{P}[\Delta^\ell h_k(X) > 0] = 1 \) is in fact too restrictive because, as has been made clear in the proof, the recurrence only implies that \( \Delta^\ell h_k(x) \) needs to be positive on \( (a+i(k), b-j(k)) \) for some indices \( i(k), j(k) \) depending on the behavior of \( X_{2k-1} \) and \( X_{2k} \) (see the line just below equation (4.16) when \( k = 1 \)). In particular the sequence necessarily stops if \( S(p) \) is bounded, since after a certain number of iterations the indicator functions defining \( \Phi_{\ell, n,j} \) will be \( 0 \) everywhere.

**Remark 4.4.** When \( g \) is a \( d \)-th degree polynomial, we obtain an exact expansion of the variance in (4.13) with respect to the \( \Gamma_k^\ell(x) \) functions \( (k = 1, \ldots, d) \) as the \( R_n^\ell \) is defined in terms of the \( n \)-th derivative of \( g \).

The functions \( \Gamma_k^\ell \) defined in (4.14) are some sort of generalization of Peccati’s and Ledoux’ gamma mentioned in the Introduction. To see this, choose \( h_k(x) = x \) for all \( k \) (arguably the most natural choice) for which \( \Delta^\ell h_k(x) = 1 \) for all \( k \) so that

\[
\Gamma_k^\ell(x) = \mathbb{E} \left[ (X_{2k} - X_{2k-1}) \Phi_{\ell,k,1}(X_1, \ldots, X_{2k-1}, x, X_{2k}, \ldots, X_2) \right] \quad (4.20)
\]

for all \( k \geq 1 \) we have. Expliciting the above leads to the expressions:

\[
\Gamma_1^\ell(x) = \mathbb{E} \left[ (X_2 - X_1) \chi_{\ell_1}(X_1, x) \chi_{-\ell_1}(x, X_2) \right] / p(x)
\]

\[
\Gamma_2^{\ell_1,\ell_2}(x) = \mathbb{E} \left[ (X_4 - X_3) \chi_{\ell_1}(X_1, X_3) \chi_{\ell_2}(X_3, x) \chi_{-\ell_2}(x, X_4) \chi_{-\ell_1}(X_4, X_2) \right] / p(x)p(X_3)p(x)p(X_4)
\]

\[
\Gamma_3^{\ell_1,\ell_2,\ell_3}(x) = \mathbb{E} \left[ (X_6 - X_5) \chi_{\ell_1}(X_1, X_3) \chi_{\ell_2}(X_3, X_5) \chi_{\ell_3}(X_5, x) \chi_{-\ell_3}(x, X_6) \chi_{-\ell_2}(X_4, X_2) \chi_{-\ell_1}(X_4, X_2) \right] / p(x)p(X_3)p(x)p(X_5)p(x)p(X_6)p(X_4)
\]

etc.
In particular $\Gamma^\ell_k(x) = \tau_p^k(x)$ is the Stein kernel of $X$. The next lemma follows by induction.

**Lemma 4.4.** If $\ell = 0$ then

$$\Gamma^0_k(x) = \frac{1}{p(x)} \mathbb{E} \left[ \frac{1}{k!(k-1)!} (x - X_1)^{k-1} (X_2 - x)^{k-1} (X_2 - X_1)_{|X_1 \leq x \leq X_2} \right].$$

(4.21)

and if $\ell \in \{-1, 1\}^k$, then

$$\Gamma^\ell_k(x) = \frac{1}{p(x)} \mathbb{E} \left[ \frac{1}{k!(k-1)!} (x - X_1)^{k-1} (X_2 - x - a\ell + 1)^{k-1} (X_2 - x - b\ell + 1)^{k-1} (X_2 - X_1)_{|X_1 + a\ell + x \leq X_2 + b\ell} \right].$$

(4.22)

where $a^{[k]}(x) = a(a + 1) \cdots (a + k - 1)$, $a\ell = \sum_{i=1}^k \frac{\ell_i (\ell_i + 1)}{2}$ and $b\ell = \sum_{i=1}^k \frac{\ell_i (1 - \ell_i)}{2}$.

**Proof.** We define $\gamma^\ell_k(x, X_1, X_2)$ such that

$$\Gamma^\ell_k(x) = \frac{1}{p(x)} \mathbb{E} \left[ \gamma^\ell_k(x, X_1, X_2) \right].$$

(4.23)

The proof is complete if we have, in the continuous case,

$$\gamma^0_k(x) = \frac{1}{k!(k-1)!} (x - X_1)^{k-1} (X_2 - x)^{k-1} (X_2 - X_1)_{|X_1 \leq x \leq X_2}$$

(4.24)

and in the discrete case,

$$\gamma^\ell_k(x) = \frac{1}{k!(k-1)!} (x - X_1 - A\ell + 1)^{k-1} (X_2 - x - B\ell + 1)^{k-1} (X_2 - X_1)_{|X_1 + A\ell \leq X_2 + B\ell}$$

(4.25)

We prove (4.24) and (4.25) by induction. Firstly, for $k = 1$, $\ell = 0, 1$ or $-1$, we obtain

$$\Gamma^\ell_k(x) = \frac{1}{p(x)} \mathbb{E} \left[ (X_2 - X_1)_{|X_1 + \ell \leq X_2 + (1 - \ell)} \right].$$

After conditioning with respect to the “extreme” variables, the expression of (4.20) can be rewritten as

$$\Gamma^\ell_{k+1}(x) = \mathbb{E} \left[ (X_2k+2 - X_{2k+1}) \Phi^\ell_{k+1,1}(X_1, \ldots, X_{2k+1}, x, X_{2k+2}, \ldots, X_2) \right]$$

$$= \mathbb{E} \left[ (X_2k+2 - X_{2k+1}) \prod_{i=1}^k \left( \frac{\chi^\ell_{k+1}(X_2i-1, X_2i+1)}{p(X_2i+1)} \right) \frac{\chi^\ell_{k+1}(X_2k+1, x) \chi^{-(k+1)}(x, X_{2k+2})}{p(X_2k+2)} \right]$$

$$= \frac{1}{p(x)} \mathbb{E}_{X_1, X_2} \left[ \mathbb{E}_{X_3, X_4} \left[ \frac{\chi^\ell_{1, k+1}(X_1, X_3)}{p(X_3)} \frac{\chi^{-(k+1)}(X_1, X_3)}{p(X_4)} \gamma^\ell_{k+1}(x, X_3, X_4) \right] X_1, X_2 \right].$$

(4.26)

This expression allows us to conclude the assertion by induction, in the continuous case and in the discrete case separately. □

**Remark 4.5.** The results e.g. from [35, 33, 19, 21, 11] permit to identify that if $X \sim p\text{ is Pearson distributed then } \Gamma^0_k(x) = \tau_p^k(x)$, and if $X \text{ is Ord distributed then } \Gamma^0_k(x) = \tau_p^k(x)$ with $f^k(x) = f(x - k + 1) \cdots f(x)$.

**Example 4.1** (Normal bounds). Direct computations show that if $X \sim N(0, 1)$ then $\Gamma^0_0(x) = 1$ and $\Gamma^1_0(x) = \frac{1}{2}$ so that the first two bounds become

$$\text{Var}[g(X)] = \mathbb{E} \left[ g'(X)^2 \right] - R_1$$

$$= \mathbb{E} \left[ g'(X)^2 \right] - \frac{1}{2} \mathbb{E} \left[ g''(X)^2 \right] + R_2$$

17
Example 4.2 (Beta bounds). Direct computations show that if $X \sim \text{Beta} (\alpha, \beta)$ then $\Gamma_1^0(x) = x(1-x)/(\alpha + \beta)$ and $\Gamma_2^0(x) = (x(1-x))^2/(2(\alpha + \beta)(\alpha + \beta + 1))$ so that the first two bounds become

$$\text{Var}[g(X)] = \frac{1}{\alpha + \beta} E \left[ X(1-X)g'(X)^2 \right] - R_1$$

$$= \frac{1}{\alpha + \beta} E \left[ X(1-X)g'(X)^2 \right] - \frac{1}{2(\alpha + \beta)(\alpha + \beta + 1)} E \left[ X^2(1-X)^2g''(X)^2 \right] + R_2$$

Example 4.3 (Gamma bounds). Direct computations show that $X \sim \text{Gamma} (\alpha, \beta)$ then $\Gamma_1^0(x) = \beta x$ and $\Gamma_2^0(x) = \frac{1}{2} \beta^2 x^2$ so that the first two bounds become

$$\text{Var}[g(X)] = \beta E \left[ Xg'(X)^2 \right] - R_1$$

$$= \beta E \left[ Xg'(X)^2 \right] - \frac{\beta^2}{2} E \left[ X^2g''(X)^2 \right] + R_2$$

Example 4.4 (Laplace bounds). Direct computations show that $X \sim \text{Laplace} (0, 1)$ then $\Gamma_1^0(x) = 1+|x|$ and $\Gamma_2^0(x) = \frac{1}{2} x^2 + |x| + 1$ so that the first two bounds become

$$\text{Var}[g(X)] = \beta E \left[ Xg'(X)^2 \right] - R_1$$

$$= \beta E \left[ Xg'(X)^2 \right] - \frac{\beta^2}{2} E \left[ X^2g''(X)^2 \right] + R_2$$

In the discrete case, there is much more flexibility in the construction of the bounds as any permutation of +1 and −1 is allowed at every stage (that is, for every k), leading to:

$$\text{Var}[g(X)] = \beta E \left[ \Gamma_1^+ (X) \Delta^− g(X) \right]^2 - R_1^+ = \beta E \left[ \Gamma_1^- (X) \Delta^+ g(X) \right]^2 - R_1^-$$

and at order 2:

$$\text{Var}[g(X)] = \beta E \left[ \Gamma_1^+ (X) \Delta^− g(X) \right]^2 - \beta E \left[ \Gamma_2^+ (X) \Delta^− g(X) \right]^2 + R_2^+$$

$$= \beta E \left[ \Gamma_1^+ (X) \Delta^− g(X) \right]^2 - \beta E \left[ \Gamma_2^- (X) \Delta^+ g(X) \right]^2 + R_2^-$$

$$\text{Var}[g(X)] = \beta E \left[ \Gamma_1^- (X) \Delta^+ g(X) \right]^2 - \beta E \left[ \Gamma_2^+ (X) \Delta^− g(X) \right]^2 + R_2^-$$

$$= \beta E \left[ \Gamma_1^- (X) \Delta^+ g(X) \right]^2 - \beta E \left[ \Gamma_2^- (X) \Delta^+ g(X) \right]^2 + R_2^-$$

where we use the concise notation $\Delta^{\ell_1,\ell_2}g(X)$ for $\Delta^{\ell_1} (\Delta^{\ell_2} g(X))$.

We detail the bounds in several settings.

Example 4.5 (Binomial bounds). Direct computations show that $X \sim \text{Bin}(n, p)$ then

$$\Gamma_1^+ (x) = (1-p)x, \quad \Gamma_1^- (x) = p(n-x)$$

and

$$\Gamma_2^+ (x) = \frac{1}{2} (1-p)^2 x(x-1), \quad \Gamma_2^- (x) = \Gamma_2^+ (x) = \frac{1}{2} p(1-p)x(n-x)$$

and

$$\Gamma_2^- (x) = \frac{1}{2} p^2 (n-x-1)(n-x)$$

so that the first bounds become at order 1:

$$\text{Var}[g(X)] = (1-p) E \left[ X(\Delta^− g(X))^2 \right] - R_1^+$$

$$= p E \left[ (n-X)(\Delta^+ g(X))^2 \right] - R_1^-$$

and at order 2:

$$\text{Var}[g(X)] = (1-p) E \left[ X(\Delta^− g(X))^2 \right] - \frac{1}{2} (1-p)^2 E \left[ X(1-X)(\Delta^− g(X))^2 \right] + R_2^+$$

$$= (1-p) E \left[ X(\Delta^− g(X))^2 \right] - \frac{1}{2} p(1-p) E \left[ X(n-X)(\Delta^− g(X))^2 \right] + R_2^-$$

$$= p E \left[ (n-X)(\Delta^+ g(X))^2 \right] - \frac{1}{2} p(1-p) E \left[ X(n-X)(\Delta^+ g(X))^2 \right] + R_2^+$$

$$= p E \left[ (n-X)(\Delta^+ g(X))^2 \right] - \frac{1}{2} p^2 E \left[ (n-X-1)(n-X)(\Delta^+ g(X))^2 \right] + R_2^-$$
We note that [38, Theorem 1.3] prove the bound

$$\text{Var}[g(X)] \leq \mathbb{E} \left( \left( \frac{X}{n} \Delta^* g(X) + \frac{n}{n} \Delta^+ g(X) \right)^2 \right)$$

which is very close to a combination of the above (see their Remark 3.3) but, as far as we can see, remains a different result. The similarity of the two results is striking but seemingly fortuitous.

Example 4.6 (Poisson bounds). Direct computations show that if $X \sim \text{Pois}(\lambda)$ then

$$\Gamma^+_1(x) = x, \quad \Gamma^-_1(x) = \lambda$$

and

$$\Gamma^{++}_2(x) = \frac{1}{2} x (x-1), \quad \Gamma^{+-}_2(x) = \Gamma^{+}_2(x) = \frac{1}{2} \lambda x \quad \text{and} \quad \Gamma^{-}_2(x) = \frac{1}{2} \lambda^2$$

so that the first bounds become at order 1:

$$\text{Var}[g(X)] = \mathbb{E} \left[ X(\Delta^- g(X))^2 \right] - R^+_1$$

$$= \lambda \mathbb{E} \left[ (\Delta^+ g(X))^2 \right] - R^-_1$$

and at order 2:

$$\text{Var}[g(X)] = \mathbb{E} \left[ X(\Delta^- g(X))^2 \right] - \frac{1}{2} \mathbb{E} \left[ X(X-1)(\Delta^- g(X))^2 \right] + R^{++}_2$$

$$= \mathbb{E} \left[ X(\Delta^- g(X))^2 \right] - \frac{1}{2} \lambda \mathbb{E} \left[ X(\Delta^+ g(X))^2 \right] + R^+_2$$

$$= \lambda \mathbb{E} \left[ (\Delta^+ g(X))^2 \right] - \frac{1}{2} \lambda^2 \mathbb{E} \left[ (\Delta^+ g(X))^2 \right] + R^-_2$$

5 Olkin-Shepp-type bounds

As mentioned in the introduction, a matrix extension of Chernoff’s gaussian bound (1.16) is due to [59]. and the result is obtained by expanding the test functions in the Hermite basis. An extension of this result to a wide class of multivariate densities is proposed in [2] (see also references therein). Once again, our notations allow to extend this result to arbitrary densities of real-valued random variables.

Theorem 5.1 (Olkin-Shepp first order bound). Let all previous notations and assumptions prevail. Then, for all $h$ such that $\Delta^{-\ell} h(X) \neq 0$ a.s.

$$\left( \begin{array}{c} \text{Var}[f(X)] \\ \text{Cov}[f(X), g(X)] \\ \text{Var}[g(X)] \end{array} \right) \leq \mathbb{E} \left[ \begin{array}{c} (\Delta^{-\ell} f(X))^2 \\ \Delta^{-\ell} f(X) \Delta^{-\ell} g(X) \\ (\Delta^{-\ell} g(X))^2 \end{array} \right] \Gamma^{\ell}_1(X) \Gamma^{\ell}_2(X)$$

with

$$\Gamma^{\ell}_1(x) = \mathbb{E} \left[ (h(X_2) - h(X_1)) \Phi^{\ell}_2(X_1, x, X_2) \right],$$

as defined in (4.14).

Remark 5.1. Taking determinants in Theorem 5.2 gives the variance-covariance inequality

$$\text{Var}[f(X)] \frac{\text{Var}[g(X)]}{\text{Cov}[f(X), g(X)]} - \{\text{Cov}[f(X), g(X)]\}^2$$

$$\leq \mathbb{E} \left[ (\Delta^{-\ell} f(X))^2 \right] \frac{\Gamma^{\ell}_1(X)}{\Delta^{-\ell} h(X)} \mathbb{E} \left[ (\Delta^{-\ell} g(X))^2 \right] \frac{\Gamma^{\ell}_2(X)}{\Delta^{-\ell} h(X)} - \mathbb{E} \left[ (\Delta^{-\ell} f(X) \Delta^{-\ell} g(X)) \right] \frac{\Gamma^{\ell}_1(X)}{\Delta^{-\ell} h(X)} \frac{\Gamma^{\ell}_2(X)}{\Delta^{-\ell} h(X)} + \frac{1}{2}.$$ (5.1)

This observation was made in [59] for the special case of $p$ being the normal distribution $\mathcal{N}(\mu, \sigma^2)$ and $h(x) = x$, for which $\Delta^{-\ell}$ is the usual derivative and $\Gamma^{\ell}_1(x) = \sigma^2$.
Remark 5.2. Multiplying the result in Theorem 5.1 by the vector \((1, 0)\)' gives the variance inequality
\[
\text{Var}[f(X)] \leq \mathbb{E} \left[ \left( \Delta^{-1} f(X) \right)^2 \Gamma_i(X) \right] \tag{5.2}
\]
which corresponds to the first order of Theorem 4.1.

The proof of Theorem 5.1 relies on a lemma which seems natural but for which we have not found a reference, and therefore we include one for completeness.

Lemma 5.1 (Matrix Cauchy-Schwarz inequality). Let \(v(x) = (a(x), b(x))' \in \mathbb{R}^2\) and \(f : \mathbb{R} \to \mathbb{R}\) any function. Then
\[
\mathbb{E} \left[ v(X) f(X) \Phi_p \ell(u, X, v) \right] \mathbb{E} \left[ v(X)' f(X) \Phi_p \ell(u, X, v) \right] \leq \mathbb{E} \left[ v(X) v(X)' \Phi_p \ell(u, X, v) \right] \mathbb{E} \left[ f(X)^2 \Phi_p \ell(u, X, v) \right], \tag{5.3}
\]
where inequality in the above is in the Loewner ordering, i.e. the difference is nonnegative definite.

Remark 5.3. In light of the proof of (5.3), there is no doubt that a similar matrix version of the Lagrange identity (4.6) holds as well. We have not explored this any further, but in particular higher order upper bound as in [20] should be within reach.

Proof of Lemma 5.1. In the proof we use the shorthand \(h\) instead of \(h(X)\), \(h_i\) instead of \(h(X_i)\), \(\Phi_p\) or \(\Phi_i\) for \(\Phi_p(u, X, X, v)\) and \(\Phi_i = \Phi_p(u, X, X, v)\). Writing (5.3) out in coordinates, it is necessary to prove that
\[
R := \begin{pmatrix}
\mathbb{E} [a^2 \Phi_p] & \mathbb{E} [f^2 \Phi_p] - \mathbb{E} [a f \Phi_p]^2 \\
\mathbb{E} [a b \Phi_p] & \mathbb{E} [b^2 \Phi_p] - \mathbb{E} [b f \Phi_p]^2
\end{pmatrix}
\]
is nonnegative definite. Next we apply (4.6) to each component of the matrix \(R\), which leads to
\[
R = \begin{pmatrix}
\mathbb{E} [(a_3 f_4 - a_4 f_3)^2 \Phi_p^{34}] & \mathbb{E} [(a_3 f_4 - a_4 f_3)(b_3 f_4 - b_4 f_3) \Phi_p^{34}] \\
\mathbb{E} [(a_3 f_4 - a_4 f_3)(b_3 f_4 - b_4 f_3) \Phi_p^{34}] & \mathbb{E} [(b_3 f_4 - b_4 f_3)^2 \Phi_p^{34}]
\end{pmatrix}.
\]
Clearly the diagonal terms are already of the correct form. For the off-diagonal terms we note that
\[
\mathbb{E}[(a_3 f_4 - a_4 f_3)(b_3 f_4 - b_4 f_3) \Phi_p^{34}] = \mathbb{E}[a_3 f_4 b_3 f_4 \Phi_p^{34}] - \mathbb{E}[a_3 f_4 b_4 f_3 \Phi_p^{34}] - \mathbb{E}[a_4 f_3 b_3 f_4 \Phi_p^{34}] + \mathbb{E}[a_4 f_3 b_4 f_3 \Phi_p^{34}]
\]
so that, using
\[
\Phi_p^{34} = \Phi_p(u, X, X, X, v) = \Phi_p(u, X, v) \Phi_p(u, X, v, v) \mathbb{I}[X_3 < X_4],
\]
we get
\[
\begin{align*}
\mathbb{E}[(a_3 f_4 - a_4 f_3)(b_3 f_4 - b_4 f_3) \Phi_p^{34}] &= \mathbb{E} [a_3 b_3 \Phi_p^{34} f_4^2 \Phi_p^{34} [X_3 < X_4]] - \mathbb{E} [a_3 b_3 \Phi_p^{34} b_4 f_4 \Phi_p^{34} [X_3 < X_4]] \\
&- \mathbb{E} [b_3 f_3 \Phi_p^{34} a_4 f_4 \Phi_p^{34} [X_3 < X_4]] + \mathbb{E} [f_3^2 \Phi_p^{34} a_4 b_4 \Phi_p^{34} [X_3 < X_4]] \\
&= \mathbb{E} [a_3 b_3 \Phi_p^{34} f_4^2 \Phi_p^{34} [X_3 < X_4]] - \mathbb{E} [a_3 f_3 \Phi_p^{34} b_4 f_4 \Phi_p^{34} [X_3 < X_4]] \\
&- \mathbb{E} [b_3 f_3 \Phi_p^{34} a_4 f_4 \Phi_p^{34} [X_3 < X_4]] + \mathbb{E} [f_3^2 \Phi_p^{34} a_4 b_4 \Phi_p^{34} [X_3 < X_4]]
\end{align*}
\]
which leads to
\[
R = \mathbb{E} \begin{pmatrix}
(a_3 f_4 - a_4 f_3) & (a_3 f_4 - a_4 f_3) & (a_3 f_4 - a_4 f_3) & (a_3 f_4 - a_4 f_3) \Phi_p^{34}
\end{pmatrix} \tag{5.4}
\]
and the claim follows.
Proof of Theorem 5.1. The variances can be expressed using (4.4) and, using successively (3.4), (4.2) and (4.3), the covariance matrix can be rewritten by

\[
\begin{pmatrix}
\text{Var}[f(X)] & \text{Cov}[f(X), g(X)] \\
\text{Cov}[f(X), g(X)] & \text{Var}[g(X)]
\end{pmatrix}
= \mathbb{E} \left[ \begin{pmatrix} f(X_2) - f(X_1) \\ g(X_2) - g(X_1) \end{pmatrix} \begin{pmatrix} f(X_2) - f(X_1) \\ g(X_2) - g(X_1) \end{pmatrix} \mathbb{1}_{X_1 < X_2} \right].
\]

Using, as usual, the representation

\[
\Delta^{-\mathbf{(Poisson bounds)}}(f(X)) \mathbb{1}_{X_1 < X_2} = \mathbb{E} \left[ \begin{pmatrix} f(X_2) - f(X_1) \\ g(X_2) - g(X_1) \end{pmatrix} \begin{pmatrix} f(X_2) - f(X_1) \\ g(X_2) - g(X_1) \end{pmatrix} \mathbb{1}_{X_1 < X_2} \right].
\]

the claim follows by applying the Matrix Cauchy-Schwarz inequality (5.3) to the vector \( v(x) = \left( \frac{\Delta^{-\mathbf{(Poisson bounds)}}(f(x))}{\sqrt{\Delta^{-\mathbf{(Poisson bounds)}}(h(x))}}, \frac{\Delta^{-\mathbf{(Poisson bounds)}}(g(x))}{\sqrt{\Delta^{-\mathbf{(Poisson bounds)}}(h(x))}} \right) \) and the function \( f(x) = \sqrt{\Delta^{-\mathbf{(Poisson bounds)}}(h(x))}. \)

Example 5.1 (Normal bounds). Direct application of (5.1) show that, by taking \( h(x) = x \), if \( X \sim \mathcal{N}(0,1) \) then

\[
\text{Var}[f(X)|\text{Var}[g(X)] - \{\text{Cov}[f(X), g(X)]\}^2 \leq \mathbb{E} \left[ (f'(X))^2 \right] \mathbb{E} \left[ (g'(X))^2 \right] - \mathbb{E} \left[ f'(X)g'(X) \right]^2
\]

A direct application of (5.2) gives us the first order upper bound

\[
\text{Var}[g(X)] \leq \mathbb{E} [g'(X)^2],
\]

which was already obtained in (1.16) and in Example 4.1.

Example 5.2 (Poisson bounds). Direct application of (5.1) show that, by taking \( h(x) = x \), if \( X \sim \text{Pois} \) then \( \text{Var}[f(X)|\text{Var}[g(X)] - \{\text{Cov}[f(X), g(X)]\}^2 \) is bounded by

\[
\mathbb{E} \left[ (\Delta^{-\mathbf{(Poisson bounds)}}(f(X))^2 X \right] \mathbb{E} \left[ (\Delta^{-\mathbf{(Poisson bounds)}}(g(X))^2 X \right] - \mathbb{E} \left[ \Delta^{-\mathbf{(Poisson bounds)}}(f(X)\Delta^{-\mathbf{(Poisson bounds)}}(g(X))X \right] ^2
\]

and by

\[
\lambda^2 \mathbb{E} \left[ (\Delta^{\mathbf{(Poisson bounds)}}(f(X))^2 X \right] \mathbb{E} \left[ (\Delta^{\mathbf{(Poisson bounds)}}(g(X))^2 X \right] - \lambda^2 \mathbb{E} \left[ \Delta^{\mathbf{(Poisson bounds)}}(f(X)\Delta^{\mathbf{(Poisson bounds)}}(g(X))X \right] ^2.
\]

Moreover, by (5.2), we find two first order upper bounds from Example 3.3:

\[
\text{Var}[g(X)] \leq \mathbb{E}[X(\Delta^{-\mathbf{(Poisson bounds)}}(g(X))^2) \text{ and Var}[g(X)] \leq \lambda \mathbb{E}[(\Delta^{\mathbf{(Poisson bounds)}}(g(X))^2)]).
\]

6 Applications of the representations to estimating Stein factors

Recall that if \( \mathcal{A}_p \), is as in (2.6) from Definition 2.4 and \( \mathcal{H} \) is a collection of \( h : \mathbb{R} \to \mathbb{R} \) belonging to \( L^1(p) \), the “\( \mathcal{A}_p \), Stein equation on \( \mathcal{H} \)” is the functional equation

\[
\mathcal{A}_p g(x) = h(x) - \mathbb{E}[h(X)], \quad x \in \mathcal{S}(p)
\]

whose solutions are

\[
g(x) = \frac{L_p^\dagger h(x)}{L_p^\dagger \eta(x)}
\]

with the convention that \( g(x) = 0 \) for all \( x \) outside of \( \mathcal{S}(p) \). Uniform in \( \mathcal{H} \) bounds on \( g \) and its derivatives are known as Stein factors; in practice it is useful to have information on \( \|g_h\|, \|\Delta^\dagger g_h\| \) also for some specific functions \( h \) (particularly \( h(x) = x \)). We conclude the paper with an application of Lemma 4.1 towards understanding properties of solutions (6.2). The result is immediate from previous developments.
Proposition 6.1. 1. If \( h \) is monotone then \( L^p \| h \| \) does not change sign.

2. If, for all \( h \in \mathcal{H} \), there exists a monotone function \( \eta \) such that \( |h(x) - h(y)| \leq k |\eta(x) - \eta(y)| \) for all \( x, y \) then the function \( g \) defined in (6.2) satisfies \( |g(x)| \leq k \). In particular if \( \mathcal{H} \) is the collection of Lipschitz functions with constant \( 1 \) then \( \|g\|_\infty \leq 1 \).

3. Let \( P \) and \( \bar{P} = 1 - P \) be the cdf and survival function of \( p \), respectively, and define the function

\[
R^\ell(x) = \frac{P(x - \ell(\ell + 1)/2)\bar{P}(x + \ell(\ell - 1)/2)}{p(x)}
\]

for all \( x \in \mathcal{S}(p) \). We always have

\[
|L^p_{\ell} h(x)| \leq 2\|h\|_{\infty} \frac{\mathbb{E}[\chi^\ell(X_1, x)]\mathbb{E}[\chi^{-\ell}(x, X_2)]}{p(x)} = 2\|h\|_{\infty} R^\ell(x)
\]

for all \( x \in \mathcal{S}(p) \).

Proof. Recall representation (6.2) which states that

\[
-L^p_{\ell} h(x) = \frac{1}{p(x)} \mathbb{E} \left[ (h(X_2) - h(X_1))\chi^\ell(X_1, x)\chi^{-\ell}(x, X_2) \right].
\]

1. If \( h(x) \) is monotone the \( h(X_2) - h(X_1) \) is of constant sign conditionally on \( \chi^\ell(X_1, x)\chi^{-\ell}(x, X_2) \).

2. Suppose that the function \( \eta \) is strictly decreasing. By definition of \( g \) we have, under the stated conditions,

\[
|g(x)| = \frac{\mathbb{E} \left[ (h(X_2) - h(X_1))\chi^\ell(X_1, x)\chi^{-\ell}(x, X_2) \right]}{\mathbb{E} \left[ (\eta(X_2) - \eta(X_1))\chi^\ell(X_1, x)\chi^{-\ell}(x, X_2) \right]}
\]

\[
\leq \frac{\mathbb{E} \left[ |h(X_2) - h(X_1)|\chi^\ell(X_1, x)\chi^{-\ell}(x, X_2) \right]}{\mathbb{E} \left[ (\eta(X_2) - \eta(X_1))\chi^\ell(X_1, x)\chi^{-\ell}(x, X_2) \right]}
\]

\[
\leq k.
\]

3. By definition,

\[
|L^p_{\ell} h(x)| = \mathbb{E} \left[ |h(X_2) - h(X_1)|\chi^\ell(X_1, x)\chi^{-\ell}(x, X_2) \right] \leq 2\|h\|_{\infty} \mathbb{E} \left[ \chi^\ell(X_1, x)\chi^{-\ell}(x, X_2) \right]
\]

which leads to the conclusion.

\[
\square
\]

Example 6.1. If \( p = \phi \) is the standard Gaussian then \( R(x) = \Phi(x)(1 - \Phi(x))/\phi(x) \) is closely related to \( r(x) \), the 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 [11 Theorem 2.3] to get

\[
\frac{1}{\sqrt{x^2 + 4 + x}} \leq \frac{1}{2} r(x) \leq R(x) \leq r(x) \leq \frac{4}{\sqrt{x^2 + 8 + 3x}}
\]

for all \( x \geq 0 \). Moreover, obviously, \( R(0) = 1/2\sqrt{\pi/2} \approx 0.626 \). If \( h \) is bounded then \( \|L^p_{\ell} h\|_{\infty} \leq \sqrt{\pi/2} \) as is well known, see e.g. [37, Theorem 3.3.1].
Figure 2: The functions $R_\ell$ for standard normal distribution (Figure 2a, along with the upper and lower bounds from [5,5]), Beta distribution with parameters 8/9 and 1/3 (Figure 2b) and Gamma distribution with parameters 1/3 and 1/3 (Figure 2c), binomial distribution with parameters (20, .2) (Figure 2d), Poisson distribution with parameter 15 (Figure 2e) and hypergeometric distribution with parameters 10, 10 and 30 (Figure 2f).

Acknowledgements

The research of YS was partially supported by the Fonds de la Recherche Scientifique – FNRS under Grant no F.4539.16. ME acknowledges partial funding via a Welcome Grant of the Université de Liège. YS also thanks Lihu Xu for organizing the “Workshop on Stein’s method and related topics” at University of Macau in December 2018, and where this contribution was first presented. GR and YS also thank Emilie Clette for fruitful discussions on a preliminary version of this work. We also thank Benjamin Arras for several pointers to relevant literature, as well as corrections on the first draft of the paper.

Finally, concerning the results in Section 6, we thank Guillaume Mijoule for communicating with us about a technique he has developed which not only leads to our bound (6.4), but also to many more higher order “Stein factors”, see [50].

References

[1] G. Afendras, N. Balakrishnan, and N. Papadatos. Orthogonal polynomials in the cumulative Ord family and its application to variance bounds. Statistics, 52(2):364–392, 2018.

[2] G. Afendras and N. Papadatos. On matrix variance inequalities. Journal of Statistical Planning and Inference, 141(11):3628–3631, 2011.

[3] G. Afendras and N. Papadatos. Strengthened Chernoff-type variance bounds. Bernoulli, 20(1):245–264, 2014.

[4] G. Afendras, N. Papadatos, and V. Papathanasiou. The discrete Mohr and Noll inequality with applications to variance bounds. Sankhyä, 69(2):162–189, 2007.

[5] G. Afendras, N. Papadatos, and V. Papathanasiou. An extended Stein-type covariance identity for the Pearson family with applications to lower variance bounds. Bernoulli, 17(2):507–529, 2011.

[6] B. Arras and C. Houdré. On Stein’s method for infinitely divisible laws with finite first moment. arXiv preprint arXiv:1712.10051, 2017.
[7] B. Arras and C. Houdré. On Stein’s method for multivariate self-decomposable laws with finite first moment. *arXiv preprint arXiv:1809.02050*, 2018.

[8] B. Arras and Y. Swan. IT formulae for gamma target: mutual information and relative entropy. *IEEE Transactions on Information Theory*, 64(2):1083–1091, 2018.

[9] A. D. Barbour and L. H. Y. Chen. *An introduction to Stein’s method*, volume 4 of Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. Singapore University Press, Singapore, 2005.

[10] A. D. Barbour and L. H. Y. Chen. *Stein’s method and applications*, volume 5 of Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. Singapore University Press, Singapore, 2005.

[11] Á. Baricz. Mills’ ratio: monotonicity patterns and functional inequalities. *Journal of Mathematical Analysis and Applications*, 340(2):1362–1370, 2008.

[12] H. J. Brascamp and E. H. Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 22(4):366–389, 1976.

[13] T. Cacoullos. On upper and lower bounds for the variance of a function of a random variable. *The Annals of Probability*, 10(3):799–809, 1982.

[14] T. Cacoullos and V. Papathanasiou. On upper and lower bounds for the variance of functions of a random variable. *Statist. Probab. Lett.*, 3:175–184, 1985.

[15] T. Cacoullos and V. Papathanasiou. Bounds for the variance of functions of random variables by orthogonal polynomials and Bhattacharyya bounds. *Statistics & Probability Letters*, 4(1):21–23, 1986.

[16] T. Cacoullos and V. Papathanasiou. Characterizations of distributions by variance bounds. *Statist. Probab. Lett.*, 7(5):351–356, 1989.

[17] T. Cacoullos and V. Papathanasiou. A generalization of covariance identity and related characterizations. *Math. Methods Statist.*, 4(1):106–113, 1995.

[18] L. H. Chen. An inequality for the multivariate normal distribution. *Journal of Multivariate Analysis*, 12(2):306–315, 1982.

[19] L. H. Chen. Poincaré-type inequalities via stochastic integrals. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 69(2):251–277, 1985.

[20] L. H. Y. Chen. Poisson approximation for dependent trials. *The Annals of Probability*, 3(3):534–545, 1975.

[21] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal approximation by Stein’s method*. Probability and its Applications (New York). Springer, Heidelberg, 2011.

[22] H. Chernoff. A note on an inequality involving the normal distribution. *The Annals of Probability*, 9(3):533–535, 1981.

[23] K. Chwialkowski, H. Strathmann, and A. Gretton. A kernel test of goodness of fit. In *International Conference on Machine Learning*, pages 2606–2615, 2016.

[24] T. A. Courtade, M. Fathi, and A. Pananjady. Existence of Stein kernels under a spectral gap, and discrepancy bound. *arXiv preprint arXiv:1703.07707*, 2017.

[25] P. Diaconis and S. Zabell. Closed form summation for classical distributions: variations on a theme of de Moivre. *Statist. Sci.*, 6(3):284–302, 1991.
[26] C. Döbler. Stein’s method of exchangeable pairs for the beta distribution and generalizations. *Electronic Journal of Probability*, 20(109):1–34, 2015.

[27] W. Ehm. Binomial approximation to the Poisson binomial distribution. *Statistics & Probability Letters*, 11(1):7–16, 1991.

[28] M. Ernst and Y. Swan. Stein based goodness-of-fit tests. In preparation, 2018.

[29] X. Fang, Q.-M. Shao, and L. Xu. Multivariate approximations in wasserstein distance by stein’s method and bismut’s formula. *Probability Theory and Related Fields*, pages 1–35, 2018.

[30] M. Fathi. Higher-order stein kernels for gaussian approximation. *arXiv preprint arXiv:1812.02703*, 2018.

[31] M. Fathi. Stein kernels and moment maps. *arXiv preprint arXiv:1804.04699*, 2018.

[32] R. E. Gaunt, A. Pickett, and G. Reinert. Chi-square approximation by Stein’s method with application to Pearson’s statistic. *Annals of Applied Probability*, 2016.

[33] L. Goldstein and G. Reinert. Stein’s method and the zero bias transformation with application to simple random sampling. *The Annals of Applied Probability*, 7(4):935–952, 1997.

[34] L. Goldstein and G. Reinert. Stein’s method for the Beta distribution and the Pólya-Eggenberger urn. *Journal of Applied Probability*, 50(4):1187–1205, 2013.

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

[36] J. Gorham and L. Mackey. Measuring sample quality with Stein’s method. In *Advances in Neural Information Processing Systems*, pages 226–234, 2015.

[37] J. Gorham and L. Mackey. Measuring sample quality with kernels. *arXiv preprint arXiv:1703.01717*, 2017.

[38] E. Hillion, O. Johnson, and Y. Yu. A natural derivative on [0, n] and a binomial Poincaré inequality. *ESAIM: Probability and Statistics*, 18:703–712, 2014.

[39] C. Houdré and A. Kagan. Variance inequalities for functions of Gaussian variables. *Journal of Theoretical Probability*, 8(1):23–30, 1995.

[40] C. Houdré and V. Pérez-Abreu. Covariance identities and inequalities for functionals on Wiener and Poisson spaces. *The Annals of Probability*, 23(1):400–419, 1995.

[41] C. Houdré, V. Pérez-Abreu, and D. Surgailis. Interpolation, correlation identities, and inequalities for infinitely divisible variables. *Journal of Fourier Analysis and Applications*, 4(6):651–668, 1998.

[42] R. W. Johnson. A note on variance bounds for a function of a Pearson variate. *Statistics & Risk Modeling*, 11(3):273–278, 1993.

[43] R. W. Johnson. A note on variance bounds for a function of a Pearson variate. *Statist. Decisions*, 11(3):273–278, 1993.

[44] S. Karlin. A general class of variance inequalities. *Multivariate Analysis: Future Directions*, Elsevier Science Publishers, New York, pages 279–294, 1993.

[45] C. A. J. Klaassen. On an inequality of Chernoff. *The Annals of Probability*, 13(3):966–974, 1985.

[46] R. Korwar. On characterizations of distributions by mean absolute deviation and variance bounds. *Annals of the Institute of Statistical Mathematics*, 43(2):287–295, 1991.
[47] M. Ledoux. L’algèbre de Lie des gradients itérés d’un générateur markovien—développements de moyennes et entropies. Ann. Sci. École Norm. Sup., 28(4):435–460, 1995.

[48] M. Ledoux, I. Nourdin, and G. Peccati. Stein’s method, logarithmic Sobolev and transport inequalities. Geometric and Functional Analysis, 25(1):256–306, 2015.

[49] C. Ley, G. Reinert, and Y. Swan. Distances between nested densities and a measure of the impact of the prior in bayesian statistics. Annals of Applied Probability, 27(1):261–291, 2017.

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

[51] C. Ley and Y. Swan. Stein’s density approach and information inequalities. Electronic Communications in Probability, 18(7):1–14, 2013.

[52] C. Ley and Y. Swan. Parametric Stein operators and variance bounds. Brazilian Journal of Probability and Statistics, 30:171–195, 2016.

[53] Q. Liu, J. Lee, and M. Jordan. A kernelized Stein discrepancy for goodness-of-fit tests. In International Conference on Machine Learning, pages 276–284, 2016.

[54] H. M. Luk. Stein’s method for the gamma distribution and related statistical applications. PhD thesis, University of Southern California, 1994.

[55] L. Mackey, J. Gorham, et al. Multivariate Stein factors for a class of strongly log-concave distributions. Electronic Communications in Probability, 21, 2016.

[56] G. Mijoule. Optimal Stein factors for continuous univariate distributions. In preparation.

[57] I. Nourdin and G. Peccati. Normal approximations with Malliavin calculus: from Stein’s method to universality. Cambridge Tracts in Mathematics. Cambridge University Press, 2012.

[58] I. Nourdin, G. Peccati, and Y. Swan. Entropy and the fourth moment phenomenon. Journal of Functional Analysis, 266:3170–3207, 2014.

[59] I. Olkin and L. Shepp. A matrix variance inequality. Journal of Statistical Planning and Inference, 130(1-2):351–358, 2005.

[60] V. Papathanasiou. Variance bounds by a generalization of the Cauchy-Schwarz inequality. Statistics & probability letters, 7(1):29–33, 1988.

[61] V. Papathanasiou. A characterization of the Pearson system of distributions and the associated orthogonal polynomials. Annals of the Institute of Statistical Mathematics, 47(1):171–176, 1995.

[62] G. Reinert. Three general approaches to Stein’s method. In An introduction to Stein’s method, volume 4. Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore, 2004.

[63] G. Reinert, G. Mijoule, and Y. Swan. Stein gradients and divergences for multivariate continuous distributions. arXiv:1806.03478, 2018.

[64] A. Saumard. Weighted Poincaré inequalities, concentration inequalities and tail bounds related to the behavior of the Stein kernel in dimension one. arXiv preprint arXiv:1804.03926, 2018.

[65] A. Saumard and J. A. Wellner. On the Isoperimetric constant, covariance inequalities and Lp-Poincaré inequalities in dimension one. arXiv preprint arXiv:1711.00668, 2017.

[66] A. Saumard and J. A. Wellner. Efron’s monotonicity property for measures on R2. Journal of Multivariate Analysis, 166:212–224, 2018.
[67] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, Vol. II: Probability theory, pages 583–602, Berkeley, Calif., 1972. Univ. California Press.

[68] C. Stein. *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986.

[69] C. Stein, P. Diaconis, S. Holmes, and G. Reinert. Use of exchangeable pairs in the analysis of simulations. In P. Diaconis and S. Holmes, editors, *Stein’s method: expository lectures and applications*, volume 46 of *IMS Lecture Notes Monogr. Ser*, pages 1–26. Beachwood, Ohio, USA: Institute of Mathematical Statistics, 2004.

[70] N. Upadhye, V. Cekanavicius, and P. Vellaisamy. On Stein operators for discrete approximations. *Bernoulli*, 23(4A):2828–2859, 2017.