ALEA, Lat. Am. J. Probab. Math. Stat. 19, 1827–1849 (2022)
DOI: 10.30757/ALEA.v19-69

On Papathanasiou's covariance expansions

Marie Ernst, Gesine Reinert and Yvik Swan

Université de Liège
E-mail address: m.ernst@uliege.be

University of Oxford
E-mail address: gesine.reinert@keble.ox.ac.uk

Université libre de Bruxelles
E-mail address: yvik.swan@ulb.be

Abstract. In this paper we provide a probabilistic representation of Lagrange's identity which we use to obtain Papathanasiou-type variance expansions of arbitrary order. Our expansions lead to generalized sequences of weights which depend on an arbitrarily chosen sequence of (non-decreasing) test functions. The expansions hold for univariate target distribution under weak assumptions, in particular they hold for continuous and lattice distributions alike. The weights are studied under different sets of assumptions either on the test functions or on the underlying distributions. Many concrete illustrations for standard probability distributions are provided (including Pearson, Ord, Laplace, Rayleigh, Cauchy, and Levy distributions).

1. Introduction

Covariances play a crucial role in probability and statistical inference, and good approximations of covariances for nonlinear functions of random variables which are based on moments of the underlying distribution of the random variables are hence sought after. In a surprisingly rarely cited paper from 1988, and using little more than the Lagrange identity (a.k.a. Cauchy-Schwarz with remainder), V. Papathanasiou proved the following variance expansion.

Theorem 1.1 (Papathanasiou's expansion, Papathanasiou (1988))**.** *Let $p(x)$ denote the density of X, an absolutely continuous real-valued random variable with finite moment of order $2n+2$ where $n \in \{0, 1, \dots\}$ is fixed. Let g be a $(n+1)$-times continuously differentiable function defined on the*

Received by the editors June 17th, 2020; accepted September 26th, 2022.
2010 *Mathematics Subject Classification.* 60E15; 26D10; 62F15.
Key words and phrases. Covariance expansion, Laplace identity, Stein's method.
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 and via the Interuniversity Attraction Pole StUDyS (IAP/P7/06). GR was supported in part by EPSRC grants EP/T018445/1 and EP/R018472/1.

support of X, and assume that $g(X)$ has finite variance. Then

$$\mathrm{Var}[g(X)] = \sum_{k=1}^{n}(-1)^{k-1}\mathbb{E}\left[(g^{(k)}(X))^2\Gamma_k(X)\right] + (-1)^n R_n \tag{1.1}$$

where R_n is a non-negative remainder term and

$$\Gamma_k(t) = \frac{(-1)^{k-1}}{k!(k-1)!}\left(\mathbb{E}\left[(X-t)^k\right]\int_{-\infty}^{t}(x-t)^{k-1}\frac{p(x)}{p(t)}dx - \mathbb{E}\left[(X-t)^{k-1}\right]\int_{-\infty}^{t}(x-t)^{k}\frac{p(x)}{p(t)}dx\right),$$

defined for all t such that $p(t) > 0$.

The weight sequence $(\Gamma_k(\cdot))_{k\geq 1}$ from (1.1) may seem unfathomable. However, in a short note from 1993, Johnson (1993) proved that if the density p is a member of the Integrated Pearson (IP) system of distributions (see Definition 3.7 below) then the weights read

$$\Gamma_k(t) = \frac{(\Gamma_1(x))^k}{k!\prod_{j=0}^{k}(1 - j\Gamma_1''(0))},$$

with Γ_1 now known to be the density's Stein covariance kernel (see Remark 3.3 for a definition and e.g. Saumard (2019); Ernst et al. (2020) for recent overviews). Many familiar univariate distributions belong to the IP system, such as the normal, beta, gamma, and Student distributions, and in all of these cases, Γ_1 is easy to compute: $\Gamma_1(x) = 1$ if X is standard Gaussian, $\Gamma_1(x) = x(1-x)/(\alpha+\beta)$ if X is beta distributed with parameters α, β, $\Gamma_1(x) = (x^2+k)/(k-1)$ if X is t_k distributed, etc. This makes (1.1) explicit in all these important cases. Another remarkable aspect of Papathanasiou's result which was first noted in Afendras et al. (2007, Theorem 3.1) is that the methods and results from Papathanasiou (1988); Johnson (1993) carry through after suitable adaptations to the discrete integer-valued (or *lattice*) setting, yielding similar flavoured expansions with weight sequences which also simplify under appropriate discretization of the Pearson assumption called Ord system of distributions (see Definition 3.11 below). This, in particular, provides infinite variance expansions for functionals of Poisson, binomial, geometric and many more distributions on the non-negative integers.

In the particular case where $X = N$ is a standard normal random variable, Papathanasiou's weight sequence simplifies to $\Gamma_k(t) = 1/k!$, leading to the variance bounds

$$\sum_{k=1}^{2n}\frac{(-1)^{k+1}}{k!}\mathbb{E}\left[g^{(k)}(N)^2\right] \leq \mathrm{Var}[g(N)] \leq \sum_{k=1}^{2n-1}\frac{(-1)^{k+1}}{k!}\mathbb{E}\left[g^{(k)}(N)^2\right] \tag{1.2}$$

for all $n \in \{1, 2, \ldots\}$. The first order (i.e. $n = 1$) term of this expansion reads as

$$\mathrm{Var}[g(N)] \leq \mathbb{E}\left[g'(X)^2\right] \tag{1.3}$$

for all differentiable functions g with finite variance. This famous Gaussian variance bound was independently discovered in Nash (1958) and Chernoff (1981) using properties of the Hermite polynomials. The Gaussian expansion (1.2) was rediscovered in Houdré and Kagan (1995). Infinite-dimensional versions of (1.2) for functionals of the Wiener and of the Poisson processes in Houdré and Pérez-Abreu (1995), concomitantly to Ledoux (1995) where properties of the Ornstein-Uhlenbeck process were used to obtain alternative expansions for general probability measures; here the terms in the expansion rest on iterations of the “carré du champ” operator. In Olkin and Shepp (2005), Chernoff's technique based on Hermite polynomials was used to extend the first order bound to the following matrix-inequality: still in the standard normal case, it holds that for all appropriate functions f and g,

$$\begin{pmatrix} \mathrm{Var}[f(N)] & \mathrm{Cov}[f(N), g(N)] \\ \mathrm{Cov}[f(N), g(N)] & \mathrm{Var}[g(N)] \end{pmatrix} \leq \begin{pmatrix} \mathbb{E}[f'(N)^2] & \mathbb{E}[f'(N)g'(N)] \\ \mathbb{E}[f'(N)g'(N)] & \mathbb{E}[g'(N)^2] \end{pmatrix} \tag{1.4}$$

where the inequality indicates that the difference is a nonnegative definite matrix. Generalizations of (1.4) can be found in Prakasa Rao (2006); Wei and Zhang (2009). In Afendras and Papadatos (2011), and using techniques which are similar to Papathanasiou's, the first order bound is extended to an expansion for distributions belonging to the Integrated Pearson (see Theorem 1 in Afendras and Papadatos (2011)) and Ord systems (see Theorem 3 in Afendras and Papadatos (2011)); the weight sequence in those expansions is the same as for the scalar expansions.

Expansions inspired by (1.2) have attracted considerable attention over the years, e.g. with extensions to matrix inequalities as in Olkin and Shepp (2005); Wei and Zhang (2009); Afendras and Papadatos (2011), to stable distributions (Koldobsky and Montgomery-Smith, 1996), to Bernoulli random vectors (Bobkov et al., 2001); more references shall be provided in the text. Aside from their intrinsic interest, they have many applications and are closely connected to a wide variety of profound mathematical questions. For statistical inference purposes, they can be used in the study of the variance of classes of estimators (see e.g. Afendras et al. (2007, section 5)), of copulas (Cuadras and Cuadras (2008)), for problems related to superconcentration (Chatterjee (2014) and Tanguy (2017)) or for the study of correlation inequalities Houdré et al. (1998) and López Blázquez and Salamanca Miño (2014). These expansions can also be interpreted as refined log-Sobolev, Poincaré or isoperimetric inequalities, see Saumard (2019). The weights appearing in the first order ($n = 1$) bounds are crucial quantities in Stein's method Fathi (2019); Ledoux et al. (2015) and their higher order extensions are closely connected to eigenvalues and eigenfunctions of certain differential operators Chen (1985).

A first connection with Stein's method was identified in Chen (1982) for multivariate Gaussian random variables, and extensions outside the Gaussian case (in a univariate setting) were obtained for instance in the papers Klaassen (1985) and Cacoullos and Papathanasiou (1997). In Ernst et al. (2020), we revisited the connection between first order (that is, $n = 1$) variance expansions and Stein's method. In particular we obtained the following first-order variance inequality (see Ernst et al. (2020, Theorem 3.5)): for any sufficiently regular function g it holds that

$$\mathrm{Var}[g(X)] \leq \mathbb{E}\left[(\Delta^{-\ell} g(X))^2 \Gamma_1^{\ell} h(X)\right] \text{ for all monotone } h \in L^1(X) \tag{1.5}$$

where $\ell = 0$ if X has a density p with interval support on the real line, in which case $\Delta^{-\ell} f(x) = f'(x)$, and $|\ell| = 1$ if X has pmf p supported on the integers in which case $\Delta^{-\ell} f(x) = (f(x+\ell) - f(x))/\ell$ (i.e. the classical finite difference operators , sometimes also called *Noerlund difference quotients*). The weight appearing in (1.5) is given by

$$\Gamma_1^{\ell} h(x) = \mathbb{E}\left[\frac{(h(X_2) - h(X_1))\mathbb{I}\left[X_1 + \mathbb{I}[\ell = 1] \leq x \leq X_2 - \mathbb{I}[\ell = -1]\right]}{p(x)\Delta^{-\ell} h(x)}\right]$$

where X_1, X_2 are iid copies of X, $\mathbb{I}[A]$ is the indicator function which is 1 if A holds, and 0 otherwise. The operator $h \mapsto {\Gamma_1}^{\ell} h$ is the "inverse density Stein operator" (see Ernst et al. (2020) for more details); the function h can be freely chosen as long as it is monotone. Taking $h(x) = x$ leads to Papathanasious's $\Gamma_1(\cdot)$ from (1.1) for continuous distributions and to the corresponding first order weight from Afendras et al. (2007) for $\mathbb{Z}$ valued distributions.

The introduction of freedom of choice through the functions h makes the inequalities widely applicable; not only does this allow for optimisation in the choice of the weights, but it also opens new angles for obtaining explicit simple weights for distributions not belonging to the Pearson or Ord systems and even provides variance expansions for functionals of distributions not admitting finite moments. Finally, we note that these generalised weights are used in Ernst and Swan (2022) in combination of Stein's method of comparison of generators to provide sharp bounds on Wasserstein, Kolmogorov and Total Variation distance between certain pairs of distributions.

From the perspective of the formalism we introduced to obtain (1.5), it seems evident that Papathanasiou's expansion (1.1), its discrete version from Afendras et al. (2007), the simplifications for IP and Ord families identified in Johnson (1993); Afendras et al. (2007), as well as the covariance matrix-inequalities from Olkin and Shepp (2005); Afendras and Papadatos (2011) all share sufficiently common traits to warrant a unified treatment within our framework. In this paper we address this task and gain new insights through the introduction of a sequence of weighting operators $h \mapsto \Gamma_k^{\ell_k} h$ which bring new interpretations to the classical weights from the literature; see our main result Theorem 2.3. Despite some notational difficulties inherent to the generality of our approach, we stress that the method of proof remains elementary; this in particular serves as a testimony to the seminality of Papathanasiou's approach in Papathanasiou (1988).

To illustrate the power of our result, we particularise our weights in a number of illustrative cases, recovering the previous literature, providing new expansions even for the already treated examples and also allowing to obtain new bounds for targets outside the reach of the current literature. Moreover we find an intriguing connection between our weighting sequences and higher order Stein operators. Elucidating the connection with infinite expansions such as Ledoux (1995), and other weighting sequences appearing within the context of Stein's method such as Azmoodeh et al. (2015), will be deferred to future work.

The paper is organised as follows. In Section 2 we provide the main results in their most abstract form. After setting up the notations (inherited mainly from Ernst et al. (2020)), Section 2.1 contains the crucial Lagrange identity (Lemma 2.2) and Section 2.2 contains the Papathanassiou-type expansion (Theorem 2.3). In Section 3 we provide illustrations by rewriting the weights appearing in Theorem 2.3 under different sets of assumptions. First, in Section 3.1 we consider a general weighting function h; next, in Section 3.2 we choose certain specific intuitively attractive h-functions (namely the identity, the cdf and the score); finally in Section 3.3 we obtain explicit expressions for various illustrative distributions (here in particular the connection with existing literature on the topic is also made). For the sake or readability, all proofs are relegated to an Appendix.

2. Infinite matrix-covariance expansions

As in Ernst et al. (2020) the setup for this paper is as follows. Let $\mathcal{X} \subset \mathbb{R}$ and equip it with some σ-algebra $\mathcal{A}$ and σ-finite measure μ. Througout this paper, $\ell \in \{-1, 0, 1\}$ though other choices are, in principle, possible. We denote $\mathrm{dom}(\Delta^\ell)$ the collection of functions $f : \mathbb{R} \to \mathbb{R}$ such that $\Delta^\ell f(x)$ exists and is finite μ-almost surely on $\mathcal{X}$. If $\ell = 0$, this corresponds to all absolutely continuous functions; if $\ell = \pm 1$ the domain is the collection of all functions on $\mathbb{Z}$.

Let X be a random variable on $\mathcal{X}$, with probability measure P^X which is absolutely continuous with respect to μ; we denote 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 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 $\overline{\mathcal{S}(p)} = [a, b]$.

As in Ernst et al. (2020) we define the generalized indicator function

$$\chi^\ell(x, y) = \mathbb{I}[x \leq y - \mathbb{I}[\ell = 1]] \tag{2.1}$$

which is defined with the obvious strict inequalities also for $x = -\infty$ and $y = \infty$, and

$$\Phi_p^\ell(u, x, v) = \chi^\ell(u, x)\chi^{-\ell}(x, v)/p(x) \tag{2.2}$$

for all $u, v \in \mathcal{S}(p)$ (note that $\Phi_p^\ell(u, x, v) = 0$ for $u > v$). Similarly we set

$$\Phi_p^\ell(u, x_1, x_2, v) = \frac{\chi^\ell(u, x_1)\chi^{|\ell|}(x_1, x_2)\chi^{-\ell}(x_2, v)}{p(x_1)p(x_2)}. \tag{2.3}$$

Moreover, for any sequence $(x_j)_{j\geq 1}$ we let $\Phi_p^{\boldsymbol{\ell},0}(x_1, x_2) = 1$ and

$$\begin{aligned}\Phi_p^{\boldsymbol{\ell},n}&(x_1, x_3, \dots, x_{2n-1}, x_{2n+1}, x_{2n+2}, x_{2n}, \dots, x_2)\\ &= \frac{1}{\prod_{i=3}^{2n+2} p(x_i)} \chi^{|\ell|}(x_{2n+1}, x_{2n+2}) \prod_{i=1}^{n} \chi^{\ell_i}(x_{2i-1}, x_{2i+1})\chi^{-\ell_i}(x_{2i+2}, x_{2i}).\end{aligned} \tag{2.4}$$

Note that $\Phi_p^{\boldsymbol{\ell},1}(x_1, x_3, x_4, x_2) = \Phi_p^\ell(x_1, x_3, x_4, x_2)$ as defined in (2.3).

Using Lemma 2.19 in Ernst et al. (2020), it is easy to see that for all x, y, it holds that

$$\chi^{|\ell|}(x, y) + \chi^{|\ell|}(y, x) = 1 + \mathbb{I}[\ell = 0]\mathbb{I}[x = y] - \mathbb{I}[\ell \neq 0]\mathbb{I}[x = y]. \tag{2.5}$$

Moreover,

$$\chi^\ell(u, y)\chi^\ell(v, y) = \chi^\ell(\max(u, v), y) \text{ and } \chi^\ell(x, u)\chi^\ell(x, v) = \chi^\ell(x, \min(u, v)). \tag{2.6}$$

We conclude with recalling Equation (28) from Ernst et al. (2020); this results motivates the covariance expansion in Theorem 2.3. 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,

$$f(x_2) - f(x_1) = \mathbb{E}\left[\Phi_p^\ell(x_1, X, x_2)\Delta^{-\ell} f(X)\right]. \tag{2.7}$$

2.1. *A probabilistic Lagrange identity.* The first ingredient for our results is the following covariance representation (all proofs are in the Appendix and it is assumed throughout that the density or pmf p satisfies Assumption A).

Lemma 2.1. *Let $X \sim p$. If X_1, X_2 are independent copies of X then for all $f, g \in L^2(p)$*

$$\begin{aligned}\mathrm{Cov}[f(X), g(X)] &= \mathbb{E}\left[\big(f(X_2) - f(X_1)\big)\big(g(X_2) - g(X_1)\big)\mathbb{I}[X_1 < X_2]\right] &(2.8)\\ &= \frac{1}{2}\mathbb{E}\left[\big(f(X_2) - f(X_1)\big)\big(g(X_2) - g(X_1)\big)\right]. &(2.9)\end{aligned}$$

Equation (2.8) is obviously not new, per se; see e.g. the variance expression in Miclo (2008, page 122). In fact, treating the lattice and continuous cases separately, one could also obtain identity (2.8) as a direct application of Lagrange's identity which reads, in the finite discrete case, as

$$\left(\sum_{k=u}^{v} a_k^2\right)\left(\sum_{k=u}^{v} b_k^2\right) - \left(\sum_{k=u}^{v} a_k b_k\right)^2 = \sum_{i=u}^{v-1}\sum_{j=i+1}^{v} (a_i b_j - a_j b_i)^2. \tag{2.10}$$

Using $a_k = g(k)\sqrt{p(k)}$ and $b_k = \sqrt{p(k)}$ for $k = 0, \dots, n$, identity (2.8) follows in the finite case. Identity (2.10) and its continuous counterpart will play a crucial role in the sequel. They are more suited to our purpose under the following form.

Lemma 2.2 (A probabilistic Lagrange identity)**.** *Fix some integer $r \in \mathbb{N}_0$ and introduce the (column) vector $\mathbf{v}(x) = (v_1(x), \cdots, v_r(x))' \in \mathbb{R}^r$. Also let $g : \mathbb{R} \to \mathbb{R}$ be any function such that $v_k g \in L^1(p)$ for all $k = 1, \dots, r$. Then*

$$\begin{aligned}&\mathbb{E}\left[\mathbf{v}(X)g(X)\Phi_p^\ell(u, X, v)\right]\mathbb{E}\left[\mathbf{v}'(X)g(X)\Phi_p^\ell(u, X, v)\right]\\ &\quad= \mathbb{E}\left[\mathbf{v}(X)\mathbf{v}'(X)\Phi_p^\ell(u, X, v)\right]\mathbb{E}\left[g^2(X)\Phi_p^\ell(u, X, v)\right] - R^\ell(u, v; \mathbf{v}, g),\end{aligned} \tag{2.11}$$

where $R^\ell(u,v;\mathbf{v},g)$ is the $r\times r$ matrix given by

$$R^\ell(u,v;\mathbf{v},g)=\mathbb{E}\left[(\mathbf{v}_3g_4-\mathbf{v}_4g_3)(\mathbf{v}_3g_4-\mathbf{v}_4g_3)'\Phi_p^\ell(u,X_3,X_4,v)\right]. \tag{2.12}$$

Here X_3,X_4 denote two independent copies of X and $\mathbf{v}_j=\mathbf{v}(X_j)$ so that $v_{ij}=v_i(X_j)$, and $g_j=g(X_j)$, $i=3,4$. When the context is clear, we abbreviate $R^\ell(u,v;\mathbf{v},g)=R(u,v)$.

2.2. *Papathanasiou-type expansion.* Now the necessary ingredients are available to give the main result of this paper. We use the notation that for a vector $\mathbf{v}=(v_1,\dots,v_r)'$ (here $'$ indicates the transpose) of functions, the operator Δ^ℓ operates on each component, so that $\Delta^\ell\mathbf{v}=(\Delta^\ell v_1,\dots,\Delta^\ell v_r)'$.

Theorem 2.3. *Fix $\ell\in\{-1,0,1\}$ and let $\boldsymbol{\ell}=(\ell_n)_{n\geq1}$ be a sequence such that $\ell_n=0$ for all n if $\ell=0$, otherwise $\ell_n\in\{-1,1\}$ arbitrarily chosen. Let $(h_n)_{n\geq1}$ be a sequence of real valued functions $h_i:\mathbb{R}\to\mathbb{R}$ such that $\mathbb{P}[\Delta^{-\ell_i}h_i(X)>0]=1$ for all $i\geq1$. Starting with some function $\mathbf{g}:\mathbb{R}\to\mathbb{R}^r$, we recursively define the sequence $(\mathbf{g}_k)_{k\geq0}$ by $\mathbf{g}_0(x)=\mathbf{g}(x)$ and $\mathbf{g}_i(x)=\Delta^{-\ell_i}\mathbf{g}_{i-1}(x)/\Delta^{-\ell_i}h_i(x)$ for all $x\in\mathcal{S}(p)$. Then, for all vectors of functions $\mathbf{f}:\mathbb{R}\to\mathbb{R}^r$ such that the expectations below exist, and all $n\geq1$, we have*

$$\mathrm{Cov}\left[\mathbf{f}(X)\right]=\sum_{k=1}^{n}(-1)^{k-1}\mathbb{E}\left[\Delta^{-\ell_k}\mathbf{f}_{k-1}(X)\Delta^{-\ell_k}\mathbf{f}_{k-1}'(X)\frac{\Gamma_p^{\boldsymbol{\ell},k}\mathbf{h}(X)}{\Delta^{-\ell_k}h_k(X)}\right]+(-1)^nR_p^{\boldsymbol{\ell},n}(\mathbf{h}) \tag{2.13}$$

where the derivatives are taken component-wise, and the weight sequences are

$$\Gamma_p^{\boldsymbol{\ell},k}\mathbf{h}(x)=\mathbb{E}\Bigg[(h_k(X_{2k})-h_k(X_{2k-1}))\Phi_p^{\ell_k}(x_{2k-1},x,x_{2k})\Phi_p^{\boldsymbol{\ell},k-1}(X_1,\dots,X_{2k-1},X_{2k},\dots,X_2)$$
$$\prod_{i=1}^{k-1}\Delta^{-\ell_i}h_i(X_{2i+1},X_{2i+2})\Bigg] \tag{2.14}$$

and

$$R_p^{\boldsymbol{\ell},n}(\mathbf{h})=\mathbb{E}\Bigg[\left(\mathbf{f}_n(X_{2n+2})-\mathbf{f}_n(X_{2n+1})\right)\left(\mathbf{f}_n(X_{2n+2})-\mathbf{f}_n(X_{2n+1})\right)'$$
$$\Phi_p^{\boldsymbol{\ell},n}(X_1,\dots X_{2n+1},X_{2n+2},\dots,X_2)\prod_{i=1}^{n}\Delta^{-\ell_i}h_i(X_{2i+1},X_{2i+2})\Bigg] \tag{2.15}$$

where $\Delta^\ell h_k(x,y)=\Delta^\ell h_k(x)\Delta^\ell h_k(y)$ and an empty product is set to 1.

Remark 2.4. If $R_p^{\boldsymbol{\ell},n}(\mathbf{h})\to0$ as $n\to\infty$ then, under the conditions of Theorem 2.3,

$$\mathrm{Cov}\left[\mathbf{f}(X)\right]=\sum_{k=1}^{\infty}(-1)^{k-1}\mathbb{E}\left[\Delta^{-\ell_k}\mathbf{f}_{k-1}(X)\Delta^{-\ell_k}\mathbf{f}_{k-1}'(X)\frac{\Gamma_p^{\boldsymbol{\ell},n}\mathbf{h}(X)}{\Delta^{-\ell_k}h_k(X)}\right]. \tag{2.16}$$

In particular when $\mathbf{f}$ is a d^{th}-degree polynomial, then $R_p^{\boldsymbol{\ell},n}(\mathbf{h})$ vanishes for $n\geq d$ and (2.13) is an exact expansion of the variance in (2.13) with respect to the $\Gamma_p^{\boldsymbol{\ell},k}\mathbf{h}(x)$ functions $(k=1,\dots,d)$.

Remark 2.5. When $\ell\neq0$ then the condition that $\mathbb{P}[\Delta^{-\ell_i}h_i(X)>0]=1$ is itself also too restrictive because, as will have been made clear in the proof (see the Appendix), the recurrence only implies that $\Delta^{-\ell_i}h_i(x)$ needs to be positive on some interval $[a+\mathbf{a}_i;b-\mathbf{b}_i]\subset[a,b]$ where $\mathbf{a}_i$ and $\mathbf{b}_i$ are positive integers (they will be properly defined in (3.4)). In particular when $\ell\neq0$ the sequence necessarily stops if $\mathcal{S}(p)$ is bounded, since after a certain number of iterations the indicator functions will be 0 everywhere.

Suppose that the remainder $R_p^{\boldsymbol{\ell},n}(\mathbf{h})$ is non negative definite. Then, taking $n=1$ in (2.13) gives an upper bound, and taking $n=2$ gives a lower bound, on the covariance, and the following holds (stated again in the case $r=2$, for the sake of clarity).

Corollary 2.6. *Let all the conditions in Theorem 2.3 prevail for $n=2$. Then*

$$\mathbb{E}\left[\Delta^{-\ell_1}f(X)\Delta^{-\ell_1}g(X)\frac{\Gamma_p^{\ell_1,1}h_1(X)}{\Delta^{-\ell_1}h_1(X)}\right]-\mathbb{E}\left[\Delta^{-\ell_2}\left(\frac{\Delta^{-\ell_1}f(X)}{\Delta^{-\ell_1}h(X)}\right)\Delta^{-\ell_2}\left(\frac{\Delta^{-\ell_1}f(X)}{\Delta^{-\ell_1}h(X)}\right)\frac{\Gamma_p^{\ell_1,\ell_2,2}(h_1,h_2)(X)}{\Delta^{-\ell_2}h_2(X)}\right]$$
$$\leq \mathrm{Cov}[f(X),g(X)]\leq \mathbb{E}\left[\Delta^{-\ell_1}f(X)\Delta^{-\ell_1}g(X)\frac{\Gamma_p^{\ell_1,1}h_1(X)}{\Delta^{-\ell_1}h_1(X)}\right].$$

Remark 2.7. When $f=g$, the upper bound for $n=1$ is a weighted Poincaré inequality of the same essence as the upper bound provided in Klaassen (1985); Cacoullos and Papathanasiou (1997) (as revisited in Ernst et al. (2020)), whereas the lower bound obtained with $n=2$ is of a different flavour.

3. About the weights in Theorem 2.3

The crucial quantities in Theorem 2.3 are the sequences of weights $\Gamma_p^{\boldsymbol{\ell},k}\mathbf{h}$ defined in (2.14). For $k=1$, the expression are straightforward to obtain (see equations (3.10) for the continuous case $\ell_1=0$ and (3.7) for the lattice case $\ell_1\in\{-1,1\}$). For larger k the situation is not so straightforward. Relevance of the higher order terms in the covariance expansions (2.13) then hinges on the tractability of these weights, which itself depends on the choice of functions $h_1,h_2,\ldots$. In this section we restrict attention to the (natural) choice $h_k(x)=h(x)$ for all k. Then, writing $\Gamma_k^{\boldsymbol{\ell}}h(x)$ instead of $\Gamma_p^{\boldsymbol{\ell},k}(h,h,\ldots)(x)$ we can express the sequence of weights as $\Gamma_k^{\boldsymbol{\ell}}h(x)=:\mathbb{E}\left[\gamma_k^{\boldsymbol{\ell}}h(X_1,x,X_2)\right]$ where, for all $k\geq 1$, we set

$$\gamma_k^{\boldsymbol{\ell}}h(x_1,x,x_2)=\mathbb{E}\Big[(h(X_{2k})-h(X_{2k-1}))\Phi_p^{\ell_k}(X_{2k-1},x,X_{2k})\Phi_p^{\boldsymbol{\ell},k-1}(x_1,X_3\ldots,X_{2k-1},X_{2k},\ldots,x_2)$$
$$\prod_{i=1}^{k-1}\Delta^{-\ell_i}h(X_{2i+1},X_{2i+2})\Big]. \tag{3.1}$$

We now study (3.1) and the resulting expressions for the weights under different sets of assumptions.

3.1. *General considerations.* The continuous case is quite easy as (2.4) simplifies when all the test functions h_i are equal and the expressions follow directly from the structure of the weight sequence, which turn out to be straightforward iterated integrals.

Lemma 3.1. *Fix $\boldsymbol{\ell}=(0,0,\ldots)$ and let h be non-decreasing. Then for all $k\geq 1$,*

$$\gamma_k^0h(x_1,x,x_2)=(h(x)-h(x_1))^{k-1}(h(x_2)-h(x))^{k-1}(h(x_2)-h(x_1))\frac{\mathbb{I}[x_1\leq x\leq x_2]}{p(x)k!(k-1)!} \tag{3.2}$$

and

$$\Gamma_k^0h(x)=\mathbb{E}\left[\big(h(x)-h(X_1)\big)^{k-1}\big(h(X_2)-h(x)\big)^{k-1}\big(h(X_2)-h(X_1)\big)\frac{\mathbb{I}[X_1\leq x\leq X_2]}{p(x)k!(k-1)!}\right]. \tag{3.3}$$

In the lattice case, simplifications of $\Gamma_k^{\boldsymbol{\ell}}h(x)$ are more difficult as (2.4) depends strongly on the chosen sequence $\boldsymbol{\ell}$. Let $\boldsymbol{\ell}=(\ell_1,\ell_2,\ldots)\in\{-1,+1\}^\infty$. For $k\geq 1, 1\leq i\leq k$, we introduce $a_i=a_i(\boldsymbol{\ell})=\mathbb{I}[\ell_i=1]$ and $b_i=b_i(\boldsymbol{\ell})=\mathbb{I}[\ell_i=-1]$ as well as

$$\mathbf{a}_k=\mathbf{a}_k(\boldsymbol{\ell})=\sum_{i=1}^k a_i \text{ and } \mathbf{b}_k=\mathbf{b}_k(\boldsymbol{\ell})=\sum_{i=1}^k b_i \tag{3.4}$$

Note that $\mathbf{a}_k$ counts the number of "+" in the first k components of $\boldsymbol{\ell}$ and $\mathbf{b}_k$ counts the corresponding number of "−", so that $\mathbf{a}_k + \mathbf{b}_k = k$. Then for (3.1), with sums over empty sets set to 1

$$\gamma_1^{\ell_1} h(x_1, x, x_2) = (h(x_2) - h(x_1)) \frac{\mathbb{I}[x_1 + a_1 \leq x \leq x_2 - b_1]}{p(x)} \tag{3.5}$$

$$\gamma_2^{\ell_1,\ell_2} h(x_1, x, x_2) = \sum_{x_3 = x_1 + a_1}^{x - a_2)} \sum_{x_4 = x + b_2}^{x_2 - b_1} (h(x_4) - h(x_3)) \Delta^{-\ell_1} h(x_3, x_4) \frac{\mathbb{I}[x_1 + \mathbf{a}_2 \leq x \leq x_2 - \mathbf{b}_2]}{p(x)} \tag{3.6}$$

and for $k \geq 3$ we have

$$\gamma_k^{\boldsymbol{\ell}} h(x_1, x, x_2) = \Bigg(\sum_{x_3 = x_1 + \mathbf{a}_{k-1}}^{x - a_k} \sum_{x_4 = x + b_k}^{x_2 - \mathbf{b}_{k-1}} (h(x_4) - h(x_3)) \Delta^{-\ell_{k-1}} h(x_3, x_4) \sum_{x_5 = x_1 + \mathbf{a}_{k-2}}^{x_3 - a_{k-1}} \sum_{x_6 = x_4 + b_{k-1}}^{x_2 - \mathbf{b}_{k-2}} \Delta^{-\ell_{k-2}} h(x_5, x_6)$$
$$\cdots \sum_{x_{2k-1} = x_1 + a_1}^{x_{2k-3} - a_2} \sum_{x_{2k+1} = x_{2k-2} + b_2}^{x_2 - b_1} \Delta^{-\ell_1} h(x_{2k-1}, x_{2k}) \Bigg) \frac{\mathbb{I}[x_1 + \mathbf{a}_k \leq x \leq x_2 - \mathbf{b}_k]}{p(x)}$$

for all $x \in \mathcal{S}(p)$ and all x_1, x_2. This is a proof of the next result.

Proposition 3.2. *Instate all previous notations. For all $k \geq 1$,*

$$\gamma_k^{\boldsymbol{\ell}} h(x_1, x, x_2) = \Bigg(\sum_{x_3 = x_1 + \mathbf{a}_{k-1}}^{x - a_k} \sum_{x_4 = x + b_k}^{x_2 - \mathbf{b}_{k-1}} (h(x_4) - h(x_3)) \psi_{k-1}^{\boldsymbol{\ell}} h(x_1, x_3, x_4, x_2) \Bigg) \frac{\mathbb{I}[x_1 + \mathbf{a}_k \leq x \leq x_2 - \mathbf{b}_k]}{p(x)}$$

where $\psi_0^{\boldsymbol{\ell}} h(x_1, x_3, x_4, x_2) = 1$ and, for $k \geq 2$, $\psi_{k-1}^{\boldsymbol{\ell}} h(x_1, x_3, x_4, x_2) = \psi_{k-1,1}^{\boldsymbol{\ell}} h(x_1, x_3) \psi_{k-1,2}^{\boldsymbol{\ell}} h(x_4, x_2)$ and

$$\psi_{k-1,1}^{\boldsymbol{\ell}} h(x_1, x_3) = \Delta^{-\ell_{k-1}} h(x_3) \sum_{x_5 = x_1 + \mathbf{a}_{k-2}}^{x_3 - a_{k-1}} \left(\Delta^{-\ell_{k-2}} h(x_5) \sum_{x_7 = x_1 + \mathbf{a}_{k-4}}^{x_5 - a_{k-2}} \left(\cdots \sum_{x_{2k-1} = x_1 + a_1}^{x_{2k-3} - a_2} \Delta^{-\ell_1} h(x_{2k-1}) \right) \right)$$

$$\psi_{k-1,2}^{\boldsymbol{\ell}} h(x_4, x_2) = \Delta^{-\ell_{k-1}} h(x_4) \sum_{x_6 = x_4 + b_{k-1}}^{x_2 - \mathbf{b}_{k-2}} \left(\Delta^{-\ell_{k-2}} h(x_6) \sum_{x_8 = x_6 + b_{k-2}}^{x_2 - \mathbf{b}_{k-3}} \left(\cdots \sum_{x_{2k} = x_{2k-2} + b_2}^{x_2 - b_1} \Delta^{-\ell_1} h(x_{2k}) \right) \right)$$

for all $x_1 + \mathbf{a}_{k-1} \leq x_3 \leq x_4 \leq x_2 - \mathbf{b}_{k-1}$.

Taking expectations in (3.5) and (3.6) we obtain

$$\Gamma_1^{\ell_1} h(x) = \frac{1}{p(x)} \mathbb{E}\left[(h(X_2) - h(X_1)) \mathbb{I}[X_1 + a_1 \leq x \leq X_2 - b_1] \right] \tag{3.7}$$

$$\Gamma_2^{\ell_1,\ell_2} h(x) = \frac{1}{p(x)} \mathbb{E}\left[\sum_{x_3 = X_1 + a_1}^{x - a_2} \sum_{x_4 = x + b_2}^{X_2 - b_1} (h(x_4) - h(x_3)) \Delta^{-\ell_1} h(x_3, x_4) \mathbb{I}[X_1 + \mathbf{a}_2 \leq x \leq X_2 - \mathbf{b}_2] \right].$$

The expressions for higher orders are easy to infer, but we have not been able to devise a formula as transparent as (3.2) for general h in the lattice case. Nevertheless, simple manageable expressions are obtainable for certain specific choices of h, particularly the case $h(x) = \mathrm{Id}(x)$, see Section 3.2.

Remark 3.3. In Ernst et al. (2020) we introduced the *inverse Stein density operator*

$$\mathcal{L}_p^{\ell} h(x) = \mathbb{E}\Big[(h(X_1) - h(X_2)) \Phi_p^{\ell}(X_1, x, X_2) \Big] \tag{3.8}$$

for $h \in L^1(p)$ and X_1, X_2 independent copies of $X \sim p$. This operator has the property of yielding solutions to so-called Stein equations, both in lattice and continuous setting; it has many important properties within the context of Stein's method. In particular it provides generalized covariance identities and, when $h(x) = \mathrm{Id}(x)$ is the identity function, it provides

$$\tau_p^\ell(x) = -\mathcal{L}_p^\ell \mathrm{Id}(x), \tag{3.9}$$

the Stein kernel of p. This function, first introduced in Stein (1986), has long been known to provide a crucial handle on the properties of p and is now studied as an object of intrinsic interest, see e.g. Courtade et al. (2019); Fathi (2019).

For absolutely continuous p, in Lemma 3.1 letting $\nu(h)$ denote the mean $\mathbb{E}[h(X)]$ we get

$$\Gamma_1^0 h(x) = \mathbb{E}\left[(h(X_2) - h(X_1))\frac{\mathbb{I}[X_1 \le x \le X_2]}{p(x)}\right] = \mathbb{E}\left[(\nu(h) - h(X))\frac{\mathbb{I}[x \le X]}{p(x)}\right] \tag{3.10}$$

which one may recognize as the inverse of the canonical Stein operator (see (3.8)); in particular taking $h(x) = \mathrm{Id}(x) = x$ the identity function, (3.10) yields the Stein kernel. There is also a connection between $\Gamma_p^{\boldsymbol{\ell},k}h$ and "higher order" Stein kernels. To see this, restrict to the continuous case $\boldsymbol{\ell} = 0$ and introduce $H_x^k(y) = (h(y) - h(x))^k/k!$. Then (3.2) becomes

$$\Gamma_k^{\mathbf{0}} h(x) = (-1)^k \left(\mathbb{E}\big[H_x^{k-1}(X)\big]\mathcal{L}_p^0 H_x^k(x) - \mathbb{E}\big[H_x^k(X)\big]\mathcal{L}_p^0 H_x^{k-1}(x)\right) \tag{3.11}$$

(see the Appendix for a proof). In the case $h(x) = x$ the expression (3.11) simplifies to Papathanasiou's weights from Theorem 1.1. This allows to make the connection between considerations related to Stein's method and the weights appearing in the expansions, as has already been observed (see e.g. Afendras et al. (2007)). Thus, our result provides a framework to the important works Papathanasiou (1988); Korwar (1991); Johnson (1993); Afendras et al. (2007, 2018), which focus on particular families of distributions, see Sections 3.3.1 and 3.3.2. Further study of this connection, in line e.g. with Fathi (2021), is outside the scope of this paper and deferred to a future publication.

3.2. *Handpicking the test functions.* We now focus on particular choices of h. The probably most intuitive choice is $h(x) = \mathrm{Id}(x)$. In this case we abbreviate $\Gamma_k^\ell \mathbf{h}(x) = \Gamma_k^\ell(x)$. If $\boldsymbol{\ell} = \mathbf{0}$ we have

$$\Gamma_k^{\mathbf{0}}(x) = \mathbb{E}\left[(X_2 - x)^{k-1}(x - X_1)^{k-1}(X_2 - X_1)\frac{\mathbb{I}[X_1 \le x \le X_2]}{k!(k-1)!p(x)}\right].$$

In the lattice case, direct computations for the first two weights give

$$\Gamma_1^{\ell_1}(x) = \mathbb{E}\left[(X_2 - X_1)\frac{\mathbb{I}[X_1 + \mathbf{a}_1 \le x \le X_2 - \mathbf{b}_1]}{p(x)}\right]$$

$$\Gamma_2^{\ell_1,\ell_2}(x) = \mathbb{E}\left[(X_2 - x - \mathbf{b}_2 + 1)(x - X_1 - \mathbf{a}_2 + 1)(X_2 - X_1)\frac{\mathbb{I}[X_1 + \mathbf{a}_2 \le x \le X_2 - \mathbf{b}_2]}{2p(x)}\right].$$

With the rising and falling factorial notation

$$f^{[k]}(x) = \prod_{j=0}^{k-1} f(x+j) \text{ and } f_{[k]}(x) = \prod_{j=0}^{k-1} f(x-j), \tag{3.12}$$

with the convention that $f^{[0]}(x) = f_{[0]}(x) = 1$, we have the following.

Lemma 3.4. *If $\boldsymbol{\ell} \in \{-1, 1\}^\infty$ then for all $k \ge 1$*

$$\Gamma_k^{\boldsymbol{\ell}}(x) = \mathbb{E}\left[(X_2 - x - \mathbf{b}_k + 1)^{[k-1]}(x - X_1 - \mathbf{a}_k + 1)^{[k-1]}(X_2 - X_1)\frac{\mathbb{I}[X_1 + \mathbf{a}_k \le x \le X_2 - \mathbf{b}_k]}{p(x)k!(k-1)!}\right]. \tag{3.13}$$

Remark 3.5. As already noted in Remark 3.3, the expression of the weights in the continuous case is already known and can be traced back to works as early as Papathanasiou (1988); the expression for the lattice case (namely equation (3.13)) is new, although a version with $\boldsymbol{\ell} = (-1, -1, -1, \ldots)$ is available from Afendras et al. (2007).

Another natural choice in the continuous case $\boldsymbol{\ell} = 0$ of an increasing function h to plug into the weights is $h(x) = P(x)$ with P the cdf of p. Then the following holds.

Lemma 3.6. *If $\boldsymbol{\ell} = 0$ and $X \sim p$ has cdf P then $\Gamma_k^0 P(x) = \frac{1}{k!(k+1)!p(x)} P(x)^k (1 - P(x))^k$.*

A final natural choice occurs whenever p is log-concave. Indeed in this case the function $h_1 = -(\log p)'$ is increasing. In particular, $\Gamma_1^0 h_1(x) = -\mathcal{L}_p^0 h_1(x) = 1$, which yields that

$$\mathrm{Cov}\left[f(X), g(X)\right] = \mathbb{E}\left[\frac{f'(X)g'(X)}{-(\log p)''(X)}\right] - R_1^0(\mathbf{h})$$

for all f, g for which the expectations exist. This expression generalizes the Brascamp-Lieb inequality from Ernst et al. (2020). For simple expressions of $R_1^0(\mathbf{h})$ one may like to choose $h_k = \mathrm{Id}$ for $k \geq 2$. This example thus benefits from the flexibility in choosing a sequence of functions $\mathbf{h}$.

3.3. *Illustrations.* The examples below take $h(x) = \mathrm{Id(x)}$ and many of them are phrased in terms of Stein kernels as in (3.9). Tables at the end of Ernst et al. (2020) give explicit expressions of Stein kernels for many standard distributions.

3.3.1. *The weights for Integrated Pearson family.* The integrated Pearson family of distribution is an important subfamily of the Pearson family as the Rodrigues polynomials form an orthogonal system for the corresponding Pearson density if and only if the density belongs to the Integrated Pearson family, see the review Afendras and Papadatos (2015).

Definition 3.7 (Integrated Pearson)**.** We say that $X \sim p$ belongs to the integrated Pearson (IP) family if X is absolutely continuous and there exist $\delta, \beta, \gamma \in \mathbb{R}$ not all equal to 0 such that $\tau_p^{\mathbf{0}}(x)\big(:= -\mathcal{L}_p^0 \mathrm{Id}(x)\big) = \delta x^2 + \beta x + \gamma$ for all $x \in \mathcal{S}(p)$.

Definition 3.7 corresponds to the continuous Pearson systems, a.k.a. integrated Pearson, as studied e.g. in Afendras and Papadatos (2014) (see their Definition 1.1). The following result holds.

Proposition 3.8. *If $X \sim p$ is IP distributed with Stein kernel $\tau_p(x) = \tau_p^{\mathbf{0}}(x) = \delta x^2 + \beta x + \gamma$ then*

$$\Gamma_k^0(x) = \frac{\tau_p(x)^k}{k! \prod_{j=0}^{k-1}(1 - j\delta)}. \tag{3.14}$$

The coefficient (δ, β, γ) of the Stein kernel can be used to directly obtain the infinite expansion of covariance for the IP family. We give the expansions for two distributions in the following examples.

Example 3.9 (Normal expansion). The standard normal distribution ϕ is an element of the IP family with $\delta = 0, \beta = 0$, and $\gamma = 1$. Direct computations show that if $X \sim \mathcal{N}(0,1)$ then $\tau_\phi(x) = 1$ so that $\Gamma_k^0(x) = \frac{1}{k!}$ for all k and for all f, g for which the expectations exist

$$\mathrm{Cov}[f(X), g(X)] = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \mathbb{E}\left[f^{(k)}(X) g^{(k)}(X)\right],$$

which extends the variance expansion (1.2) to a covariance expansion.

Example 3.10 (Beta expansion). The $\mathrm{Beta}(a,b)$ distribution is an element of the IP family with $\delta = -\frac{1}{a+b}, \beta = \frac{1}{a+b}$, and $\gamma = 0$; then $\tau_{\mathrm{Beta(a,b)}}(x) = \frac{x(1-x)}{a+b}$. If $X \sim \mathrm{Beta}(a,b)$ then $\Gamma_k^0(x) = (x(1-x))^k / (k!(a+b)^{[k]})$ for $k \geq 1$, so that for all f, g for which the expectations exist

$$\mathrm{Cov}[f(X), g(X)] = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!(a+b)^{[k]}} \mathbb{E}\left[f^{(k)}(X) g^{(k)}(X) X^k (1-X)^k\right].$$

3.3.2. *The weights for Cumulative Ord family.* In this subsubsection the superscript + denotes $\ell = 1$ and the superscript − denotes $\ell = -1$. We use the same notations as in Afendras et al. (2007).

Definition 3.11 (Cumulative Ord families)**.** We say that $X \sim p$ belongs to the cumulative Ord (CO) family if X is discrete and there exist $\delta, \beta, \gamma \in \mathbb{R}$ not all equal to 0 such that $\tau_p^-(x)\big(:= -\mathcal{L}_p^-(\mathrm{Id})\big) = \delta x^2 + \beta x + \gamma$ for all $x \in \mathcal{S}(p)$.

Note that it follows that for this distribution p, $\tau_p^+(x) = \frac{p(x-1)}{p(x)}\tau_p^-(x-1) = x(\delta x + \beta + 1)$. The following result holds.

Proposition 3.12. *If $X \sim p$ is CO distributed with $\tau_p^-(x) = \delta x^2 + \beta x + \gamma$ then*

$$\Gamma_k^{\boldsymbol{\ell}}(x) = \frac{1}{k! \prod_{j=0}^{k-1}(1 - j\delta)} \left(\tau_p^+(x)\right)_{[\mathbf{a}_k]} \left(\tau_p^-(x)\right)^{[\mathbf{b}_k]}. \tag{3.15}$$

Remark 3.13. By taking only k forward differences, i.e., $\boldsymbol{\ell} = (-1, \dots, -1)$, we deduce the result of Afendras et al. (2007, Theorem 4.1). In particular, their Table 1 illustrates the expression of $\Gamma_k^{\boldsymbol{\ell}}(x)$ for some discrete distributions from the CO family.

In the lattice case, there is much more flexibility in the construction of the bounds as any permutation of +1 and −1 is allowed for every k, leading to:

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

and for an order 2 expansion, for any of the four choices of $(\ell_1, \ell_2) \in \{-1, +1\}^2$,

$$\mathrm{Var}[g(X)] = \mathbb{E}\left[\Gamma_1^{\ell_1}(X)(\Delta^{-\ell_1} g(X))^2\right] - \mathbb{E}\left[\Gamma_2^{\ell_1,\ell_2}(X)(\Delta^{-\ell_1,-\ell_2} g(X))^2\right] + R_2^{\ell_1,\ell_2}$$

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

Example 3.14 (Binomial expansion). The Binomial(n, θ) distribution is an element of the CO family with $\delta = 0, \beta = -\theta$, and $\gamma = n\theta$; its Stein kernels are $\tau^-(x) = \theta(n - x)$ and $\tau^+(x) = (1 - \theta)x$. Hence

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

so that the order 1 expansions are

$$\begin{aligned} \mathrm{Var}[g(X)] &= (1 - \theta)\mathbb{E}\left[X(\Delta^- g(X))^2\right] - R_1^+ &(3.16)\\ &= \theta\mathbb{E}\left[(n - X)(\Delta^+ g(X))^2\right] - R_1^-; &(3.17)\end{aligned}$$

choosing a linear combination of (3.16) and (3.17) with weights θ and $1 - \theta$, respectively, yields

$$\mathrm{Var}[g(X)] = n\theta(1 - \theta)\mathbb{E}\left[\frac{X}{n}(\Delta^- g(X))^2 + \frac{n - X}{n}(\Delta^+ g(X))^2\right] - \theta R_1^+ - (1 - \theta)R_1^-. \tag{3.18}$$

In Hillion et al. (2014, Theorem 1.3) the "natural binomial derivative" $\nabla_n g(x) = \frac{x}{n}\Delta^- g(x) + \frac{n-x}{n}\Delta^+ g(x)$ is introduced and used to prove the Poincaré inequality

$$\mathrm{Var}[g(X)] \le n\theta(1 - \theta)\mathbb{E}\left[\left(\nabla_n g(X)\right)^2\right].$$

The connection with (3.18) is easy to see because (see e.g. Hillion et al. (2014, Remark 3.3))

$$\left(\nabla_n g(x)\right)^2 = \frac{x}{n}(\Delta^- g(x))^2 + \frac{n - x}{n}(\Delta^+ g(x))^2 - \frac{x(n - x)}{n^2}(\Delta^{+-} g(x))^2.$$

Moving to the second order, direct computations show that

$$\Gamma_2^{+,+}(x) = \frac{1}{2}(1 - \theta)^2 x(x - 1)\mathbb{I}[1 \le x \le n], \quad \Gamma_2^{+,-}(x) = \Gamma_2^{-,+}(x) = \frac{1}{2}\theta(1 - \theta)x(n - x)\mathbb{I}[0 \le x \le n]$$

and $\Gamma_2^{-,-}(x) = \frac{1}{2}\theta^2(n - x)(n - x - 1)\mathbb{I}[0 \le x \le n - 1]$

leading to four order 2 expansions for $\mathrm{Var}[g(X)]$. These yield for example the lower variance bound

$$\mathrm{Var}[g(X)] \geq n\theta(1-\theta)\left\{\mathbb{E}\left[\left(\nabla_n g(X)\right)^2\right] - \frac{n-2}{2}\mathbb{E}\left[\frac{X(n-X)}{n^2}(\Delta^{+-}g(X))^2\right]\right\}.$$

Combining these inequalities yields that for $0<\theta<1$,

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

3.3.3. *Examples which are not integrated Pearson or cumulative Ord distributions.*

Example 3.15 (Laplace expansion). If $X \sim \mathrm{Laplace}(0,1)$ (i.e. $p(x) = e^{-|x|}/2$ on $\mathbb{R}$) then $\Gamma_1^0(x) = 1+|x|$ and $\Gamma_2^0(x) = \frac{1}{2}x^2 + |x| + 1$ so that the first two terms in the variance expansion are

$$\begin{aligned}\mathrm{Var}[g(X)] &= \mathbb{E}\left[(1+|X|)g'(X)^2\right] - R_1 \\ &= \mathbb{E}\left[(1+|X|)g'(X)^2\right] - \mathbb{E}\left[(1+|X|+X^2/2)g''(X)^2\right] + R_2.\end{aligned}$$

The general expression for Γ_k is quite simple:

$$\Gamma_k^0(x) = \sum_{j=0}^k \frac{|x|^j}{j!}.$$

The structure of this sequence seems to indicate that this distribution is of a different nature than IP distributions; this is also illustrated in the properties of the corresponding Stein operator (which is best described as a second order differential operator), see Eichelsbacher and Thäle (2015); Pike and Ren (2014).

Example 3.16 (Rayleigh expansion). If $X \sim \mathrm{Rayleigh}(0,1)$ (i.e. $p(x) = xe^{-x^2/2}$ on $\mathbb{R}^+$) then $\tau_p^0(x)$ does not take on an agreeable form. Nevertheless the choice $h(x) = x^2$ leads to

$$\frac{\Gamma_k^0 h(x)}{h'(x)} = \frac{2^{k-2}}{k!}x^{2(k-1)}.$$

Example 3.17 (Cauchy expansion). The standard Cauchy distribution lacks moments; nevertheless taking $h(x) = \arctan(x)$ leads to

$$\frac{\Gamma_k^0(x)}{h'(x)} = \frac{1}{4^k(k+1)!(k)!}(1+x^2)^2\left(\pi^2 - 4\arctan(x)^2\right)^k.$$

Example 3.18 (Levy expansion). The *pdf* of the standard Levy distribution is given by $\frac{1}{\sqrt{2\pi}}e^{\frac{1}{2x}}x^{-\frac{3}{2}}$. Similarly as in the previous example, taking $h(x) = P(x)$,

$$\frac{\Gamma_k^0(x)}{h'(x)} = \binom{k+1}{2}\frac{1}{k!(k+1)!}\pi e^{1/x}x^3\big((1-P(x))P(x)\big)^k.$$

3.4. *Final remarks.* As already mentioned in the Introduction, it would be interesting to provide a spectral or orthogonal decomposition lens to our weighting sequence; for instance direct comparison with the weights identified e.g. in Ledoux (1995) may be possible (recall for instance that in the Beta case of Example 3.10 Jacobi polynomials are eigenvectors of the natural diffusion operator). We also conjecture that there is a Stein's method interpretation of these higher order weighting operators $h \mapsto \Gamma_k^{\ell} h$, and in particular in the context of Stein-Malliavin calculus where such higher order operators already exist, see Azmoodeh et al. (2015) (from which the notation Γ for the weights is inspired). Tackling this problem exceeds the scope of the present article.

Appendix A. Proofs

Proof of Lemma 2.1: The equivalence between (2.9) and (2.8) follows from the fact that $\mathbb{I}[X_1 < X_2] + \mathbb{I}[X_1 = X_2] + \mathbb{I}[X_1 > X_2] = 1$ and

$$\mathbb{E}\left[\big(f(X_2)-f(X_1)\big)\big(g(X_2)-g(X_1)\big)\mathbb{I}[X_1<X_2]\right] = \mathbb{E}\left[\big(f(X_2)-f(X_1)\big)\big(g(X_2)-g(X_1)\big)\mathbb{I}[X_2<X_1]\right].$$

Without loss of generality in (2.9) it can be assumed that $\mathbb{E}[f(X)] = \mathbb{E}[g(X)] = 0$. Evaluating the expectation (2.9) through expanding the product yields the assertion. □

Proof of Lemma 2.2: First, from (2.6) in it follows directly that

$$\Phi_p^\ell(u,x_1,x_2,v)\mathbb{I}[x_1 \neq x_2] = \mathbb{I}[x_1 \neq x_2]\chi^{|\ell|}(x_1,x_2)\Phi_p^\ell(u,x_1,v)\Phi_p^\ell(u,x_2,v). \tag{A.1}$$

With the abbreviations as introduced in the statement of the lemma, the (i,j) entry of the $r \times r$ matrix $R(u,v)$ is

$$\begin{aligned}(R(u,v))_{i,j} &:= \mathbb{E}\left[(v_{i3}g_4 - v_{i4}g_3)(v_{j3}g_4 - v_{j4}g_3)\Phi_p^\ell(u,X_3,X_4,v)\right]\\ &= \mathbb{E}\left[\mathbb{I}[X_3 \neq X_4](v_{i3}g_4 - v_{i4}g_3)(v_{j3}g_4 - v_{j4}g_3)\chi^{|\ell|}(X_3,X_4)\Phi_p^\ell(u,X_3,v)\Phi_p^\ell(u,X_4,v)\right],\end{aligned}$$

where we used (A.1) in the last step. Next, using (2.5), $\mathbb{I}[x_1 \neq x_2](\chi^{|\ell|}(x_1,x_2) + \chi^{|\ell|}(x_2,x_1)) = \mathbb{I}[x_1 \neq x_2]$ and by symmetry,

$$\begin{aligned}&\mathbb{E}\left[\mathbb{I}[X_3 \neq X_4](v_{i3}g_4 - v_{i4}g_3)(v_{j3}g_4 - v_{j4}g_3)\chi^{|\ell|}(X_3,X_4)\Phi_p^\ell(u,X_3,v)\Phi_p^\ell(u,X_4,v)\right]\\ &\quad= \mathbb{E}\left[\mathbb{I}[X_4 \neq X_3](v_{i3}g_4 - v_{i4}g_3)(v_{j3}g_4 - v_{j4}g_3)\chi^{|\ell|}(X_4,X_3)\Phi_p^\ell(u,X_3,v)\Phi_p^\ell(u,X_4,v)\right].\end{aligned}$$

Thus

$$\begin{aligned}2(R(u,v))_{i,j} &= \mathbb{E}\left[\mathbb{I}[X_3 \neq X_4](v_{i3}g_4 - v_{i4}g_3)(v_{j3}g_4 - v_{j4}f_3)\chi^{|\ell|}(X_3,X_4)\Phi_p^\ell(u,X_3,v)\Phi_p^\ell(u,X_4,v)\right]\\ &\quad+\mathbb{E}\left[\mathbb{I}[X_4 \neq X_3](v_{i3}g_4 - v_{i4}g_3)(v_{j3}g_4 - v_{j4}g_3)\chi^{|\ell|}(X_4,X_3)\Phi_p^\ell(u,X_3,v)\Phi_p^\ell(u,X_4,v)\right]\\ &= \mathbb{E}\left[\mathbb{I}[X_3 \neq X_4](v_{i3}g_4 - v_{i4}g_3)(v_{j3}g_4 - v_{j4}g_3)\Phi_p^\ell(u,X_3,v)\Phi_p^\ell(u,X_4,v)\right]\\ &= \mathbb{E}\left[(v_{i3}g_4 - v_{i4}g_3)(v_{j3}g_4 - v_{j4}g_3)\Phi_p^\ell(u,X_3,v)\Phi_p^\ell(u,X_4,v)\right].\end{aligned}$$

Now we exploit the independence of X_3 and X_4 to obtain

$$\begin{aligned}2(R(u,v))_{i,j} = 2\mathbb{E}&\left[v_{i3}v_{j3}\Phi_p^\ell(u,X_3,v)\right]\mathbb{E}\left[g_4^2\Phi_p^\ell(u,X_4,v)\right]\\ &- 2\mathbb{E}\left[v_{i3}g_3\Phi_p^\ell(u,X_3,v)\right]\mathbb{E}\left[v_{j4}g_4\Phi_p^\ell(u,X_4,v)\right].\end{aligned}$$

The assertion follows by dividing by 2 and re-arranging the equation. □

Proof of Theorem 2.3: First by direct verification we note that the following recursion for $\Phi_p^{\boldsymbol{\ell},n}$ holds. Starting from $\Phi_p^{\boldsymbol{\ell},1}(x_1,x_3,x_4,x_2) = \Phi_p^{\ell_1}(x_1,x_3,x_4,x_2)$ we have for $n \geq 2$

$$\begin{aligned}&\Phi_p^{\boldsymbol{\ell},n}(x_1,x_3,\dots,x_{2n-1},x_{2n+1},x_{2n+2},x_{2n},\dots,x_2)\\ &\quad= \Phi_p^{\ell_n}(x_{2n-1},x_{2n+1},x_{2n+2},x_{2n})\Phi_p^{\boldsymbol{\ell},n-1}(x_1,x_3,\dots,x_{2n-1},x_{2n},\dots,x_2)\end{aligned} \tag{A.2}$$

for any sequence $(x_j)_{j\geq 1}$. We abbreviate

$$\Phi_{n,1}^{\boldsymbol{\ell}}(x_1,x_3,\dots,x_{2n-1},x,x_{2n},\dots,x_2) = \Phi_p^{\ell_n}(x_{2n-1},x,x_{2n})\Phi_p^{\boldsymbol{\ell},n-1}(x_1,x_3,\dots,x_{2n-1},x_{2n},\dots,x_2). \tag{A.3}$$

The proof uses induction in n. First consider $n = 1$. Let X_1, X_2, X_3, X_4 be independent copies of X. Starting from (2.8),

$$\begin{aligned}
&\mathrm{Cov}\,[\mathbf{f}(X)] = \mathbb{E}[(\mathbf{f}(X_2) - \mathbf{f}(X_1))(\mathbf{f}(X_2) - \mathbf{f}(X_1))'\mathbb{I}[X_1 < X_2]] \\
&= \mathbb{E}\left[\mathbb{E}\left[\Phi_p^{\ell_1}(X_1, X_3, X_2)\Delta^{-\ell_1}\mathbf{f}(X_3) \,|\, X_1, X_2\right]\mathbb{E}\left[\Phi_p^{\ell_1}(X_1, X_4, X_2)\Delta^{-\ell_1}\mathbf{f}(X_4) \,|\, X_1, X_2\right]' \mathbb{I}[X_1 < X_2]\right]
\end{aligned}$$

where we used (2.7) in the last step. Now for any h_1 such that $\mathbb{P}[\Delta^{-\ell_1}h_1(X) > 0] = 1$, dividing and multiplying by $\sqrt{\Delta^{-\ell_1}h_1(X)}$ and applying Lemma 2.2 (Lagrange identity) with

$$\mathbf{v}(x) = \frac{\Delta^{-\ell_1}\mathbf{f}(x)}{\sqrt{\Delta^{-\ell_1}h_1(x)}} \quad \text{and } g(x) = \sqrt{\Delta^{-\ell_1}h_1(x)} \tag{A.4}$$

gives

$$\begin{aligned}
&\mathrm{Cov}\,[\mathbf{f}(X)] + \mathbb{E}\left[R^{\ell_1}(X_1, X_2; \mathbf{v}, g)\mathbb{I}[X_1 < X_2]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbf{v}(X)\mathbf{v}'(X)\Phi_p^{\ell_1}(X_1, X, X_2) \,|\, X_1, X_2\right]\mathbb{E}\left[g^2(X)\Phi_p^{\ell_1}(X_1, X, X_2) \,|\, X_1, X_2\right]\mathbb{I}[X_1 < X_2]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{\Delta^{-\ell_1}\mathbf{f}(X)\Delta^{-\ell_1}\mathbf{f}'(X)}{\Delta^{-\ell_1}h_1(X)}\Phi_p^{\ell_1}(X_1, X, X_2)|X_1, X_2\right]\mathbb{E}\left[\Delta^{-\ell_1}h_1(X)\Phi_p^{\ell_1}(X_1, X, X_2)|X_1, X_2\right]\mathbb{I}[X_1 < X_2]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{\Delta^{-\ell_1}\mathbf{f}(X)\Delta^{-\ell_1}\mathbf{f}'(X)}{\Delta^{-\ell_1}h_1(X)}\Phi_p^{\ell_1}(X_1, X, X_2)|X_1, X_2\right](h_1(X_2) - h(X_1))\mathbb{I}[X_1 < X_2]\right] \qquad \text{(A.5)}
\end{aligned}$$

with the last equality following from (2.7). Note that, in the lattice case, the strict inequality in the indicator $\mathbb{I}[X_1 < X_2]$ is implicit in $\Phi_p^{\ell_1}(X_1, X, X_2) = \chi^{\ell_1}(X_1, X)\chi^{-\ell_1}(X, X_2)/p(X)$ (and hence a fortiori also in $\Phi_p^{\ell_1}(X_1, X_3, X_4, X_2)$; in the continuous case there is no difference between $\mathbb{I}[X_1 < X_2]$ and $\mathbb{I}[X_1 \leq X_2]$. Hence unconditioning yields

$$\begin{aligned}
&\mathbb{E}\left[\frac{\Delta^{-\ell_1}\mathbf{f}(X)\Delta^{-\ell_1}\mathbf{f}'(X)}{\Delta^{-\ell_1}h_1(X)}\Phi_p^{\ell_1}(X_1, X, X_2)(h_1(X_2) - h_1(X_1))\mathbb{I}[X_1 < X_2]\right] \\
&\quad = \mathbb{E}\left[\frac{\Delta^{-\ell_1}\mathbf{f}(X)\Delta^{-\ell_1}\mathbf{f}'(X)}{\Delta^{-\ell_1}h_1(X)}\Phi_p^{\ell_1}(X_1, X, X_2)(h_1(X_2) - h_1(X_1))\right] \\
&\quad = \mathbb{E}\left[\Delta^{-\ell_1}\mathbf{f}(X)\Delta^{-\ell_1}\mathbf{f}'(X)\frac{\Gamma_p^{\ell_1,1}h_1(X)}{\Delta^{-\ell_1}h_1(X)}\right],
\end{aligned}$$

giving the first term in the covariance expansion (2.13). With the notation (A.4), the remainder term in (A.5) is

$$\begin{aligned}
&\mathbb{E}\left[R^{\ell_1}(X_1, X_2; \mathbf{v}, g)\mathbb{I}[X_1 < X_2]\right] \\
&\quad = \mathbb{E}\left[\mathbb{E}\left[(\mathbf{v}_3g_4 - \mathbf{v}_4g_3)(\mathbf{v}_3g_4 - \mathbf{v}_4g_3)'\Phi_p^{\ell_1}(X_1, X_3, X_4, X_2)|X_1, X_2\right]\mathbb{I}[X_1 < X_2]\right].
\end{aligned}$$

Now,

$$\mathbf{v}_3g_4 = \frac{\Delta^{-\ell_1}\mathbf{f}(X_3)}{\sqrt{\Delta^{-\ell_1}h_1(X_3)}}\sqrt{\Delta^{-\ell_1}h_1(X_4)} = \frac{\Delta^{-\ell_1}\mathbf{f}(X_3)}{\Delta^{-\ell_1}h_1(X_3)}\sqrt{\Delta^{-\ell_1}h_1(X_3)\Delta^{-\ell_1}h_1(X_4)}$$

and $\sqrt{\Delta^{-\ell_1} h_1(X_3)\Delta^{-\ell_1} h_1(X_4)}$ is a common factor, so that

$$\begin{aligned}
&\mathbb{E}\left[R^{\ell_1}(X_1, X_2; \mathbf{v}, g)\mathbb{I}[X_1 < X_2]\right] \\
&= \mathbb{E}\Bigg[\left(\frac{\Delta^{-\ell_1}\mathbf{f}(X_3)}{\Delta^{-\ell_1}h_1(X_3)} - \frac{\Delta^{-\ell_1}\mathbf{f}(X_4)}{\Delta^{-\ell_1}h_1(X_4)}\right)\left(\frac{\Delta^{-\ell_1}\mathbf{f}(X_3)}{\Delta^{-\ell_1}h_1(X_3)} - \frac{\Delta^{-\ell_1}\mathbf{f}(X_4)}{\Delta^{-\ell_1}h_1(X_4)}\right)' \\
&\qquad \times \left(\sqrt{\Delta^{-\ell_1}h_1(X_3)\Delta^{-\ell_1}h_1(X_4)}\right)^2 \Phi_p^{\ell_1}(X_1, X_3, X_4, X_2)\mathbb{I}[X_1 < X_2]\Bigg] \\
&= \mathbb{E}\left[(\mathbf{f}_1(X_3) - \mathbf{f}_1(X_4))(\mathbf{f}_1(X_3) - \mathbf{f}_1(X_4))'\Delta^{-\ell_1}h_1(X_3)\Delta^{-\ell_1}h_1(X_4)\Phi_p^{\ell_1}(X_1, X_3, X_4, X_2)\right] \\
&= R_p^{\ell_1,1}(\mathbf{h})
\end{aligned}$$

as required; here $\mathbf{h} = h_1$. Thus the assertion holds for $n = 1$.

To obtain the complete claim, we proceed by induction and suppose that the claim holds at some n. It remains to show that

$$R_p^{\boldsymbol{\ell},n}(\mathbf{h}) = \mathbb{E}\left[\Delta^{-\ell_{n+1}}\mathbf{f}_n(X)\Delta^{-\ell_{n+1}}\mathbf{f}_n'(X)\frac{\Gamma_p^{\boldsymbol{\ell},n+1}\mathbf{h}(X)}{\Delta^{-\ell_{n+1}}h_{n+1}(X)}\right] - R_p^{\boldsymbol{\ell},n+1}(\mathbf{h}). \tag{A.6}$$

To this purpose, starting from (2.15), we simply apply the same process as above: for $x_{2n+1} < x_{2n+2}$, we use

$$\mathbf{f}_n(x_{2n+2}) - \mathbf{f}_n(x_{2n+1}) = \mathbb{E}\left[\Delta^{-\ell_{n+1}}\mathbf{f}_n(X)\Phi_p^{\ell_{n+1}}(x_{2n+1}, X, x_{2n+2})\right]$$

as well as the Lagrange identity (2.11) and simple conditioning to obtain that

$$\begin{aligned}
R_p^{\boldsymbol{\ell},n}(\mathbf{h}) &= \mathbb{E}\Bigg[(\mathbf{f}_n(X_{2n+2}) - \mathbf{f}_n(X_{2n+1}))(\mathbf{f}_n(X_{2n+2}) - \mathbf{f}_n(X_{2n+1}))' \\
&\qquad \Phi_p^{\boldsymbol{\ell},n}(X_1, \dots X_{2n+1}, X_{2n+2}, \dots, X_2)\prod_{i=1}^n \Delta^{-\ell_i}h_i(X_{2i+1}, X_{2i+2})\Bigg] \\
&= \mathbb{E}\Bigg[\mathbb{E}\left[\Delta^{-\ell_{n+1}}\mathbf{f}_n(X_{2n+3})\Phi_p^{\ell_{n+1}}(X_{2n+1}, X_{2n+3}, X_{2n+2})|X_{2n+1}, X_{2n+2}\right] \\
&\qquad \mathbb{E}\left[\Delta^{-\ell_{n+1}}\mathbf{f}_n'(X_{2n+4})\Phi_p^{\ell_{n+1}}(X_{2n+1}, X_{2n+4}, X_{2n+2})|X_{2n+1}, X_{2n+2}\right] \\
&\qquad \Phi_p^{\boldsymbol{\ell},n}(X_1, \dots X_{2n+1}, X_{2n+2}, \dots, X_2)\prod_{i=1}^n \Delta^{-\ell_i}h_i(X_{2i+1}, X_{2i+2})\Bigg].
\end{aligned}$$

Now for any h_{n+1} such that $\mathbb{P}[\Delta^{-\ell_{n+1}}h_{n+1}(X) > 0] = 1$, dividing and multiplying by $\sqrt{\Delta^{-\ell_{n+1}}h_{n+1}(X)}$ and applying Lemma 2.2 with

$$\mathbf{v}_{n+1}(x) = \frac{\Delta^{-\ell_{n+1}}\mathbf{f}_n(x)}{\sqrt{\Delta^{-\ell_{n+1}}h_{n+1}(x)}} \quad \text{and } g_{n+1}(x) = \sqrt{\Delta^{-\ell_{n+1}}h_{n+1}(x)} \tag{A.7}$$

we obtain with (2.14)

$$
\begin{aligned}
& R_p^{\boldsymbol{\ell},n}(\mathbf{h}) - \mathbb{E}\Big[\mathbb{E}\left[R^{\ell_{n+1}}(X_{2n+1}, X_{2n+2}; \mathbf{v_{n+1}}, g_{n+1})|X_{2n+1}, X_{2n+2}\right] \\
& \qquad \mathbb{I}[X_{2n+1} < X_{2n+2}]\Phi_p^{\boldsymbol{\ell},n}(X_1, \dots X_{2n+1}, X_{2n+2}, \dots, X_2) \prod_{i=1}^n \Delta^{-\ell_i} h_i(X_{2+1}, X_{2i+2})\Big] \\
= {} & \mathbb{E}\Big[\mathbb{E}\left[\mathbf{v}_{n+1}(X)\mathbf{v}'_{n+1}(X)\Phi_p^{\ell_{n+1}}(X_{2n+1}, X, X_{2n+2})|X_{2n+1}, X_{2n+2}\right] \\
& \quad \times\mathbb{E}\left[g_{n+1}^2(X)\Phi_p^{\ell_{n+1}}(X_{2n+1}, X, X_{2n+2})|X_{2n+1}, X_{2n+2}\right] \\
& \quad \mathbb{I}[X_{2n+1} < X_{2n+2}]\Phi_p^{\boldsymbol{\ell},n}(X_1, \dots X_{2n+1}, X_{2n+2}, \dots, X_2) \prod_{i=1}^n \Delta^{-\ell_i} h_i(X_{2+1}, X_{2i+2})\Big] \\
= {} & \mathbb{E}\Big[\mathbb{E}\left[\mathbf{v}_{n+1}(X)\mathbf{v}'_{n+1}(X)\Phi_p^{\ell_{n+1}}(X_{2n+1}, X, X_{2n+2})\right](h_{n+1}(X_{2n+2}) - h_{n+1}(X_{2n+1})) \\
& \Phi_p^{\boldsymbol{\ell},n}(X_1, \dots X_{2n+1}, X_{2n+2}, \dots, X_2) \prod_{i=1}^n \Delta^{-\ell_i} h_i(X_{2i+1}, X_{2i+2})\Big] \\
= {} & \mathbb{E}\left[\Delta^{-\ell_{n+1}}\mathbf{f}_n(X)\Delta^{-\ell_{n+1}}\mathbf{f}'_n(X)\frac{\Gamma_p^{\boldsymbol{\ell},n+1}\mathbf{h}(X)}{\Delta^{-\ell_{n+1}}h_{n+1}(X)}\right] \qquad \text{(A.8)}
\end{aligned}
$$

where we used (A.7) in the last step. Thus we have recovered the first summand in (A.6). For the remainder term in (A.8), leaving out the negative sign, the notation (A.7) gives

$$
\begin{aligned}
& \mathbb{E}\big[\mathbb{E}\left[R^{\ell_{n+1}}(X_{2n+1}, X_{2n+2}; \mathbf{v_{n+1}}, g_{n+1})|X_{2n+1}, X_{2n+2}\right] \\
& \qquad \mathbb{I}[X_{2n+1} < X_{2n+2}]\Phi_p^{\boldsymbol{\ell},n}(X_1, \dots X_{2n+1}, X_{2n+2}, \dots, X_2) \prod_{i=1}^n \Delta^{-\ell_i} h_i(X_{2i+1}, X_{2i+2})\big] \\
= {} & \mathbb{E}\Big[(\mathbf{v}_{n+1,2n+3}g_{n+1,2n+4} - \mathbf{v}_{n+1,2n+4}g_{n+1,2n+3})(\mathbf{v}_{n+1,2n+3}g_{n+1,2n+4} - \mathbf{v}_{n+1,2n+4}g_{n+1,2n+3})' \\
& \quad \Phi_p^{\ell_{n+1}}(X_{2n+1}, X_{2n+3}, X_{2n+4}, X_{2n+2})\Phi_p^{\boldsymbol{\ell},n}(X_1, \dots X_{2n+1}, X_{2n+2}, \dots, X_2) \\
& \quad \prod_{i=1}^n \Delta^{-\ell_i} h_i(X_{2i+1}, X_{2i+2})\Big]
\end{aligned}
$$

and again extracting the common factor $\sqrt{\Delta^{-\ell_{n+1}}h_{n+1}(X_{2n+3})\Delta^{-\ell_{n+1}}h_{n+1}(X_{2n+4})}$ and re-arranging yields the assertion. □

Proof of Lemma 3.1: Let $x_1 \le x \le x_2$ and h an increasing function. Direct application of the definitions with (2.4) lead to

$$
\begin{aligned}
p(x)\gamma_k^0 h(x_1, x, x_2) = \int_{x_1}^{x}\int_{x}^{x_2}\int_{x_3}^{x}\int_{x}^{x_4}\cdots\int_{x_{2k-3}}^{x}\int_{x}^{x_{2k-2}} (h(x_{2k}) - h(x_{2k-1})h'(x_{2k-1})h'(x_{2k})\mathrm{d}x_{2k}\mathrm{d}x_{2k-1} \\
\cdots h'(x_5)h'(x_6)\mathrm{d}x_6\mathrm{d}x_5 h'(x_3)h'(x_4)\mathrm{d}x_4\mathrm{d}x_3.
\end{aligned}
$$

Applying the change of variables $u_k = h(x_k), k = 1, \dots, 2k$ and setting $u = h(x)$ we see that the sequence $\gamma_k^0 h$ depends only on the iterated integrals

$$
\iota_k(u_1, u, u_2) := \int_{u_1}^{u}\int_{u}^{u_2}\int_{u_3}^{u}\int_{u}^{u_4}\cdots\int_{u_{2k-3}}^{u}\int_{u}^{u_{2k-2}} (u_{2k} - u_{2k-1})\mathrm{d}u_{2k}\mathrm{d}u_{2k-1}\cdots\mathrm{d}u_6\mathrm{d}u_5\mathrm{d}u_4\mathrm{d}u_3
$$

which we can write recursively as

$$\iota_1(u_1,u,u_2) = u_2 - u_1$$
$$\iota_k(u_1,u,u_2) = \int_{u_1}^{u}\int_{u}^{u_2} \iota_{k-1}(u_3,u,u_4)\mathrm{d}u_4\mathrm{d}u_3, \qquad k \geq 2.$$

It remains to show that

$$\iota_k(u_1,u,u_2) = (u_2-u)^{k-1}(u-u_1)^{k-1}(u_2-u_1)\frac{\mathbb{I}[u_1 \leq u \leq u_2]}{k!(k-1)!} \tag{A.9}$$

for all $k \geq 1$. We proceed by induction on k. Clearly $\iota_1(u_1,u,u_2) = (u_2-u_1)\mathbb{I}[u_1 \leq u \leq u_2]$, as required. Next suppose that (A.9) holds. Then

$$\begin{aligned}\iota_{k+1}(u_1,u,u_2) &= \frac{1}{k!(k-1)!}\int_{u_1}^{u}\int_{u}^{u_2}(u_4-u)^{k-1}(u-u_3)^{k-1}(u_4-u_3)\mathrm{d}u_4\mathrm{d}u_3 \\ &= \frac{1}{k!(k-1)!}\int_{u_1}^{u}\int_{u}^{u_2}(u_4-u)^{k}(u-u_3)^{k-1}\mathrm{d}u_4\mathrm{d}u_3 \\ &\quad + \frac{1}{k!(k-1)!}\int_{u_1}^{u}\int_{u}^{u_2}(u_4-u)^{k}(u-u_3)^{k-1}\mathrm{d}u_4\mathrm{d}u_3 \\ &= \frac{(u_2-u)^{k+1}(u-u_1)^k + (u_2-u)^k(u-u_1)^{k+1}}{(k+1)!k!}\end{aligned}$$

which leads to the claim. □

Proof of Identity (3.11)*:* Identity (3.11) follows from Lemma 3.1 by using $h(X_2)-h(X_1) = h(X_2)-h(x)+h(x)-h(X_1)$ and $\mathbb{I}[X_1 \leq x \leq X_2]\mathbb{I}[X_1 \neq X_2] = \mathbb{I}[X_1 \leq x]\mathbb{I}[X_2 \geq x]\mathbb{I}[X_1 \neq X_2]$ to get

$$\begin{aligned}\Gamma_k^{\mathbf{0}}h(x) &= (-1)^{k-1}\frac{1}{p(x)}\mathbb{E}\big[H_x^{k-1}(X)\mathbb{I}[X \leq x]\big]\mathbb{E}\big[H_x^{k}(X)\mathbb{I}[X \geq x]\big] \\ &\quad + (-1)^k\frac{1}{p(x)}\mathbb{E}\big[H_x^{k}(X)\mathbb{I}[X \leq x]\big]\mathbb{E}\big[H_x^{k-1}(X)\mathbb{I}[X \geq x]\big] \\ &= (-1)^{k-1}\mathbb{E}\big[H_x^{k-1}(X)\big]\frac{1}{p(x)}\mathbb{E}\big[H_x^{k}(X)\mathbb{I}[X \geq x]\big] + (-1)^k\mathbb{E}\big[H_x^{k}(X)\big]\frac{1}{p(x)}\mathbb{E}\big[H_x^{k-1}(X)\mathbb{I}[X \geq x]\big]\end{aligned} \tag{A.10}$$

where the last equality follows from

$$\mathbb{E}\big[H_x^k(X)\big] = \mathbb{E}\big[H_x^k(X)\mathbb{I}[X \leq x]\big] + \mathbb{E}\big[H_x^k(X)\mathbb{I}[X \geq x]\big].$$

Upon noting that

$$\begin{aligned}&-\mathcal{L}_p^0 H_x^k(x) \\ &= \frac{1}{p(x)}\Big\{\mathbb{E}\big[H_x^k(X_2)\mathbb{I}[X_1 < x < X_2]\big] - \mathbb{E}\big[H_x^k(X_1)\mathbb{I}[X_1 < x < X_2]\big]\Big\} \\ &= \frac{1}{p(x)}\Big\{\mathbb{E}\big[H_x^k(X_2)\mathbb{I}[x < X_2]\big]\mathbb{P}[x > X_1] - \mathbb{E}\big[H_x^k(X_1)\mathbb{I}[X_1 < x]\mathbb{P}[x < X_2]\big]\Big\} \\ &= \frac{1}{p(x)}\Big\{\mathbb{E}\big[H_x^k(X_2)\mathbb{I}[x < X_2]\big] - \mathbb{E}\big[H_x^k(X_2)\mathbb{I}[x < X_2]\big]\mathbb{P}[x < X_1] \\ &\qquad\qquad - \mathbb{E}\big[H_x^k(X_1)\mathbb{I}[X_1 < x]\mathbb{P}[x < X_2]\big]\Big\}\end{aligned}$$

with $P(x) = \mathbb{P}[X \leq x]$ we obtain

$$\frac{1}{p(x)}\mathbb{E}\big[H_x^k(X)\mathbb{I}[X \geq x]\big] = -\mathcal{L}_p^0 H_x^k(x) + \frac{1-P(x)}{p(x)}\mathbb{E}\big[H_x^k(X)\big],$$

the required result is obtained after straightforward simplifications by writing

$$\begin{aligned}\Gamma_k^{\mathbf{0}} h(x) &= (-1)^{k-1}\mathbb{E}\big[H_x^{k-1}(X)\big]\frac{1}{p(x)}\mathbb{E}\big[H_x^k(X)\mathbb{I}[X\ge x]\big] + (-1)^k\mathbb{E}\big[H_x^k(X)\big]\frac{1}{p(x)}\mathbb{E}\big[H_x^{k-1}(X)\mathbb{I}[X\ge x]\big]\\ &= (-1)^{k-1}\Big(-\mathbb{E}\big[H_x^{k-1}(X)\big]\mathcal{L}_p^0 H_x^k(x) + \mathbb{E}\big[H_x^k(X)\big]\mathcal{L}_p^0 H_x^{k-1}(x)\Big)\\ &\quad + (-1)^{k-1}\frac{1-P(x)}{p(x)}\Big(\mathbb{E}\big[H_x^{k-1}(X)\big]\mathbb{E}\big[H_x^k(X)\big] - \mathbb{E}\big[H_x^k(X)\big]\mathbb{E}\big[H_x^{k-1}(X)\big]\Big)\end{aligned}$$

and noticing that the last term cancels. □

Proof of Lemma 3.4: We shall prove that

$$\gamma_k^{\boldsymbol{\ell}}(x_1,x,x_2) := \gamma_k^{\boldsymbol{\ell}}\mathrm{Id}(x_1,x,x_2) = (x_2-x)_{\{k-1;\boldsymbol{\ell}\}}(x-x_1)^{\{k-1;\boldsymbol{\ell}\}}(x_2-x_1)\frac{\mathbb{I}[x_1+\mathbf{a}_k \le x \le x_2-\mathbf{b}_k]}{p(x)k!(k-1)!}. \tag{A.11}$$

The claim is obvious from (3.2) in the continuous case. For the lattice case, the assertion is proved by induction in k; the cases $k=1$ and $k=2$ need to be asserted to start the induction. The case $k=1$ is immediate. For $k=2$, we show that

$$\gamma_2^{\ell_1,\ell_2}(X_1,x,X_2) = \frac{1}{2}(x-X_1-a_{\boldsymbol{\ell}}(2)+1)(X_2-x-b_{\boldsymbol{\ell}}(2)+1)(X_2-X_1)\frac{\mathbb{I}[X_1+a_{\boldsymbol{\ell}}(2)\le x\le X_2-b_{\boldsymbol{\ell}}(2)]}{p(x)}$$

for $\ell_i \in \{-1,1\}$. To this end, from Proposition 3.2 where we sum over (x_3,x_4) instead of (y,z), we obtain

$$\begin{aligned}\gamma_2^{\ell_1,\ell_2}(x_1,x,x_2) &= \sum_{x_3=x_1+a_1}^{x-a_2}\sum_{x_4=x+b_2}^{x_2-b_1}(x_4-x_3)\frac{\mathbb{I}[x_1+\mathbf{a}_2\le x\le x_2-\mathbf{b}_2]}{p(x)}\\ &= \frac{1}{2}(x-x_1-\mathbf{a}_2+1)(x_2-x-\mathbf{b}_2+1)(x_2-x_1)\frac{\mathbb{I}[x_1+\mathbf{a}_2\le x\le x_2-\mathbf{b}_2]}{p(x)}\end{aligned}$$

as required.

To conclude the argument, we prove the identity (A.11) by induction: we suppose the claims hold for k and investigate its validity for $k+1$. The definition of $\Gamma_p^{\boldsymbol{\ell},k}$ in (2.14) gives

$$\gamma_{k+1}^{\boldsymbol{\ell}}(x_1,x,x_2) = \mathbb{E}\left[\frac{\chi^{\ell_1}(x_1,X_3)}{p(X_3)}\frac{\chi^{-\ell_1}(X_4,x_2)}{p(X_4)}\gamma_k^{\ell_2,\dots,\ell_{k+1}}(X_3,x,X_4)\right] \tag{A.12}$$

Now we can plug-in the induction assumption (A.11) into (A.12):

$$\begin{aligned}&\gamma_{k+1}^{\boldsymbol{\ell}}(x_1,x,x_2)\\ &= \mathbb{E}\Bigg[(X_4-x-\mathbf{b}'_k+1)^{[k-1]}(x-X_3-\mathbf{a}'_k+1)^{[k-1]}(X_4-X_3)\frac{\mathbb{I}[X_3+\mathbf{a}'_k\le x\le X_4-\mathbf{b}'_k]}{p(x)k!(k-1)!}\\ &\qquad \frac{\chi^{\ell_1}(x_1,X_3)}{p(X_3)}\frac{\chi^{-\ell_1}(X_4,x_2)}{p(X_4)}\Bigg]\\ &= \sum_{x_3=x_1+a_1}^{x-\mathbf{a}'_k}\sum_{x_4=x+\mathbf{b}'_k}^{x_2-b_1}(x_4-x-\mathbf{b}'_k+1)^{[k-1]}(x-x_3-\mathbf{a}'_k+1)^{[k-1]}(x_4-x_3)\frac{\mathbb{I}[x_1+\mathbf{a}_{k+1}\le x\le x_2-\mathbf{b}_{k+1}]}{p(x)}\\ &= (x_2-x-\mathbf{b}_{k+1}+1)^{[k]}(x-x_1-\mathbf{a}_{k+1}+1)^{[k]}(x_2-x_1)\frac{\mathbb{I}[x_1+\mathbf{a}_{k+1}\le x\le x_2-\mathbf{b}_{k+1}]}{p(x)}\end{aligned}$$

where $\mathbf{a}'_k = \sum_{i=2}^{k+1} a_i$ and $\mathbf{b}'_k = \sum_{i=2}^{k+1} b_i$. □

Proof of Lemma 3.6: By Lemma 3.1 and (A.10), we have

$$\Gamma_k^0 P(x) = \frac{1}{p(x)k!(k-1)!}\mathbb{E}\left[(P(x)-P(X_1))^{k-1}\mathbb{I}[X_1 \le x]\right]\mathbb{E}\left[(P(X_2)-P(x))^{k}\mathbb{I}[X_2 \ge x]\right] + \frac{1}{p(x)k!(k-1)!}\mathbb{E}\left[(P(x)-P(X_1))^{k}\mathbb{I}[X_1 \le x]\right]\mathbb{E}\left[(P(X_2)-P(x))^{k-1}\mathbb{I}[X_2 \ge x]\right].$$

Moreover, using integration by substitution,

$$\mathbb{E}\left[(P(x)-P(X_1))^{k}\mathbb{I}[X_1 \le x]\right] = \int_a^x (P(x)-P(x_1))^k p(x_1)dx_1 = -\int_{P(x)}^0 u^k du = \frac{P(x)^{k+1}}{k+1}$$

$$\mathbb{E}\left[(P(X_2)-P(x))^{k}\mathbb{I}[X_2 \ge x]\right] = \int_x^b (P(x_2)-P(x))^k p(x_2)dx_2 = \int_0^{1-P(x)} u^k du = \frac{(1-P(x))^{k+1}}{k+1},$$

and the conclusion follows. □

Proof of Proposition 3.8: This argument is inspired by Johnson (1993, Theorem 2). By Lemma 3.1,

$$\gamma_k^0(x_1,x,x_2) = (x-x_1)^{k-1}(x_2-x)^{k-1}(x_2-x_1)\frac{\mathbb{I}[x_1 \le x \le x_2]}{p(x)k!(k-1)!} = (x-x_1)^{k-1}(x_2-x)^{k-1}(x_2-\mu+\mu-x_1)\frac{\mathbb{I}[x_1 \le x]\mathbb{I}[x \le x_2]}{p(x)k!(k-1)!}$$

Therefore, $\Gamma_k^0(x)$ can be decomposed using expectations:

$$\Gamma_k^0(x) = \frac{1}{p(x)k!(k-1)!}\Bigg(\mathbb{E}\left[(x-X_1)^{k-1}\mathbb{I}[X_1 \le x]\right]\mathbb{E}\left[(X_2-\mu)(X_2-x)^{k-1}\mathbb{I}[x \le X_2]\right] + \mathbb{E}\left[(\mu-X_1)(x-X_1)^{k-1}\mathbb{I}[X_1 \le x]\right]\mathbb{E}\left[(X_2-x)^{k-1}\mathbb{I}[x \le X_2]\right]\Bigg). \tag{A.13}$$

In the continuous setting, the Stein kernel τ_p is such that is satisfies for $X \sim p$ with mean μ and differentiable f such that the expectations exist,

$$\mathbb{E}[(X-\mu)f(X)] = \mathbb{E}[\tau_p(X)f'(X)].$$

Integrating by parts we thus obtain

$$\mathbb{E}\left[(X_2-\mu)(X_2-x)^{k-1}\mathbb{I}[X_2 \ge x]\right] = \mathbb{E}\left[\tau_p(X_2)(k-1)(X_2-x)^{k-2}\mathbb{I}[X_2 \ge x]\right]$$

and

$$\mathbb{E}\left[(\mu-X_1)(x-X_1)^{k-1}\mathbb{I}[X_1 \le x]\right] = \mathbb{E}\left[\tau_p(X_1)(k-1)(x-X_1)^{k-2}\mathbb{I}[X_1 \le x]\right].$$

When we plug it into (A.13), we get

$$\Gamma_k^0(x) = \frac{k-1}{p(x)k!(k-1)!}\mathbb{E}\Big[(x-X_1)^{k-2}(X_2-x)^{k-2}(\tau_p(X_2)(x-X_1)+\tau_P(X_1)(X_2-x))\mathbb{I}[X_1 \le x \le X_2]\Big].$$

Using the particular form of τ_p for the IP family, Taylor expansion of $\tau_p(x)$ around x_1 and around x_2 and using that $\tau_p''(x) = 2\delta$ is constant gives

$$(x-x_1)\tau_p(x_2) + (x_2-x)\tau_p(x_1) = \tau_p(x)(x_2-x_1) + \frac{\tau_p''(x)}{2}(x-x_1)(x_2-x)(x_2-x_1)$$

Therefore,

$$\begin{aligned}\Gamma_k^0(x) =&\frac{k-1}{k!(k-1)!}\frac{1}{p(x)}\mathbb{E}\Big[(x-X_1)^{k-2}(X_2-x)^{k-2}\mathbb{I}[X_1\le x\le X_2]\\&\left(\tau_p(x)(X_2-X_1)+\frac{\tau_p''(x)}{2}(x-X_1)(X_2-x)(X_2-X_1)\right)\Big]\\=&\frac{\tau_p(x)}{k}\Gamma_{k-1}^0(x)+\frac{\tau_p''(x)(k-1)}{2}\Gamma_k^0(x)\\=&\frac{1}{k\left(1-\frac{k-1}{2}\tau_p''(x)\right)}\tau_p(x)\Gamma_{k-1}^0(x)\end{aligned}$$

The assertion follows from iterating this expression and using $\Gamma_1^0(x)=\tau_p(x)$ and $\tau_p''(x)=2\delta$. □

Proof of Proposition 3.12: By induction, we only have to prove the relation with respect to ℓ_{k+1}, i.e.,

$$\Gamma_{k+1}^{\boldsymbol{\ell},1}(x)=\frac{\tau_p^+(x-\mathbf{a}_k)}{(k+1)(1-k\delta)}\Gamma_k^{\boldsymbol{\ell}}(x) \text{ and } \Gamma_{k+1}^{\boldsymbol{\ell},-1}(x)=\frac{\tau_p^-(x+\mathbf{b}_k)}{(k+1)(1-k\delta)}\Gamma_k^{\boldsymbol{\ell}}(x).$$

The following argument is inspired by Afendras et al. (2007). Using (A.11) and a similar proof as in the IP case (Proposition 3.8), we may rewrite $\Gamma_{k+1}^{\boldsymbol{\ell},1}(x)$ using expectations:

$$\begin{aligned}\Gamma_{k+1}^{\boldsymbol{\ell},1}(x)=\frac{1}{p(x)}\frac{1}{k!(k+1)!}\Big(&\\&\mathbb{E}\left[(x-X_1-\mathbf{a}_k)^{[k]}\mathbb{I}[X_1+\mathbf{a}_k+1\le x]\right]\mathbb{E}\left[(X_2-\mu)(X_2-x-\mathbf{b}_k+1)^{[k]}\mathbb{I}[x\le X_2-\mathbf{b}_k]\right]\\&+\mathbb{E}\left[(\mu-X_1)(x-X_1-\mathbf{a}_k)^{[k]}\mathbb{I}[X_1+\mathbf{a}_k+1\le x]\right]\mathbb{E}\left[(X_2-x-\mathbf{b}_k+1)^{[k]}\mathbb{I}[x\le X_2-\mathbf{b}_k]\right]\Big).\end{aligned} \tag{A.14}$$

With the notation (3.12) is it straightforward to verify that for all x we have

$$\Delta^\ell\left(f^{[k]}(x)\right)=f^{[k-1]}(x+a_\ell)\sum_{j=0}^{k-1}\Delta^\ell f(x+j). \tag{A.15}$$

In particular, for all x,a, we have

$$\begin{aligned}\Delta^-\left((x-a+1)^{[k]}\mathbb{I}[x\ge a]\right)&=k(x-a+1)^{[k-1]}\mathbb{I}[x\ge a]\\\Delta^+\left((a+1-x)^{[k]}\mathbb{I}[x\le a]\right)&=-k(a+1-x)^{[k-1]}\mathbb{I}[x\le a]\\\Delta^-\left((a-x)^{[k]}\mathbb{I}[x<a]\right)&=-k(a-x+1)^{[k-1]}\mathbb{I}[x\le a]\end{aligned}$$

The Stein kernel τ_p^ℓ for discrete distributions satisfies for $X\sim p$ with mean μ and functions f such that the expectations exist,

$$\mathbb{E}[(X-\mu)f(X)]=\mathbb{E}[\tau_p^\ell(X)\Delta^{-\ell}f(X-\ell)],$$

see for example Ley et al. (2017). Hence, with (A.15), we may use the discrete integration by parts formula to rewrite

$$\mathbb{E}\left[(X_2-\mu)(X_2-x-\mathbf{b}_k+1)^{[k]}\mathbb{I}[x\le X_2-\mathbf{b}_k]\right]=k\mathbb{E}\left[\tau_p^+(X_2)(X_2-x-\mathbf{b}_k+1)^{[k-1]}\mathbb{I}[x\le X_2-\mathbf{b}_k]\right]$$

and

$$\mathbb{E}\left[(\mu - X_1)(x - X_1 - \mathbf{a}_k)^{[k]}\mathbb{I}[X_1 \le x - \mathbf{a}_k - 1]\right] = \mathbb{E}\left[(\mu - X_1)(x - X_1 - \mathbf{a}_k)^{[k]}\mathbb{I}[X_1 \le x - \mathbf{a}_k]\right]$$
$$= k\mathbb{E}\left[\tau_p^+(X_1)(x - X_1 - \mathbf{a}_k + 1)^{[k-1]}\mathbb{I}[X_1 \le x - \mathbf{a}_k]\right].$$

After plugging these equations into (A.14) and some further algebraic developments, we obtain

$$\Gamma_{k+1}^{\boldsymbol{\ell},1}(x) = \frac{1}{p(x)}\frac{1}{k!(k+1)!}\Bigg(k\tau_p^+(x - \mathbf{a}_k)$$
$$\mathbb{E}\left[(x - X_1 - \mathbf{a}_k + 1)^{[k-1]}(X_2 - x - \mathbf{b}_k + 1)^{[k-1]}(X_2 - X_1)\mathbb{I}[X_1 + \mathbf{a}_k \le x \le X_2 - \mathbf{b}_k]\right]$$
$$+ \delta k\mathbb{E}\left[(X_2 - X_1)(x - X_1 - \mathbf{a}_k)(X_2 - x + k - \mathbf{b}_k)\mathbb{I}[X_1 + \mathbf{a}_k + 1 \le x \le X_2 - \mathbf{b}_k]\right]\Bigg)$$
$$= \frac{\tau_p^+(x - \mathbf{a}_k)}{k+1}\Gamma_k^{\boldsymbol{\ell}}(x) + \delta k \Gamma_{k+1}^{\boldsymbol{\ell},1}(x)$$

which gives the assertion. The equivalent result for $\Gamma_{k+1}^{\boldsymbol{\ell},-1}(x)$ can easily be obtained. □

Acknowledgements

YS thanks Lihu Xu for organizing the "Workshop on Stein's method and related topics" at University of Macau in December 2018, and where an early version of this contribution was first presented. GR and YS thank Emilie Clette for fruitful discussions on a preliminary version of this work. YS thanks Jean-Pierre Schneiders for discussions on the weights. We also thank Benjamin Arras for several pointers to relevant literature, as well as suggestions on the first draft of the paper. All three authors would like to thank the associate editor and the anonymous referees for their reports, which have lead to an improvement of the paper.

References

Afendras, G., Balakrishnan, N., and Papadatos, N. Orthogonal polynomials in the cumulative Ord family and its application to variance bounds. *Statistics*, **52** (2), 364–392 (2018). MR3772186.

Afendras, G. and Papadatos, N. On matrix variance inequalities. *J. Statist. Plann. Inference*, **141** (11), 3628–3631 (2011). MR2817368.

Afendras, G. and Papadatos, N. Strengthened Chernoff-type variance bounds. *Bernoulli*, **20** (1), 245–264 (2014). MR3160581.

Afendras, G. and Papadatos, N. Integrated Pearson family and orthogonality of the Rodrigues polynomials: a review including new results and an alternative classification of the Pearson system. *Appl. Math. (Warsaw)*, **42** (2-3), 231–267 (2015). MR3422115.

Afendras, G., Papadatos, N., and Papathanasiou, V. The discrete Mohr and Noll inequality with applications to variance bounds. *Sankhyā*, **69** (2), 162–189 (2007). MR2428867.

Azmoodeh, E., Peccati, G., and Poly, G. Convergence towards linear combinations of chi-squared random variables: a Malliavin-based approach. In *In memoriam Marc Yor—Séminaire de Probabilités XLVII*, volume 2137 of *Lecture Notes in Math.*, pp. 339–367. Springer, Cham (2015). MR3444306.

Bobkov, S. G., Götze, F., and Houdré, C. On Gaussian and Bernoulli covariance representations. *Bernoulli*, **7** (3), 439–451 (2001). MR1836739.

Cacoullos, T. and Papathanasiou, V. Characterizations of distributions by generalizations of variance bounds and simple proofs of the CLT. *J. Statist. Plann. Inference*, **63** (2), 157–171 (1997). MR1491576.

Chatterjee, S. *Superconcentration and related topics*. Springer Monographs in Mathematics. Springer, Cham (2014). ISBN 978-3-319-03885-8; 978-3-319-03886-5. MR3157205.

Chen, L. H. Y. An inequality for the multivariate normal distribution. *J. Multivariate Anal.*, **12** (2), 306–315 (1982). MR661566.

Chen, L. H. Y. Poincaré-type inequalities via stochastic integrals. *Z. Wahrsch. Verw. Gebiete*, **69** (2), 251–277 (1985). MR779459.

Chernoff, H. A note on an inequality involving the normal distribution. *Ann. Probab.*, **9** (3), 533–535 (1981). MR614640.

Courtade, T. A., Fathi, M., and Pananjady, A. Existence of Stein kernels under a spectral gap, and discrepancy bounds. *Ann. Inst. Henri Poincaré Probab. Stat.*, **55** (2), 777–790 (2019). MR3949953.

Cuadras, C. M. and Cuadras, D. Eigenanalysis on a bivariate covariance kernel. *J. Multivariate Anal.*, **99** (10), 2497–2507 (2008). MR2463403.

Eichelsbacher, P. and Thäle, C. Malliavin-Stein method for variance-gamma approximation on Wiener space. *Electron. J. Probab.*, **20**, Paper No. 123, 28 (2015). MR3425543.

Ernst, M., Reinert, G., and Swan, Y. First-order covariance inequalities via Stein's method. *Bernoulli*, **26** (3), 2051–2081 (2020). MR4091101.

Ernst, M. and Swan, Y. Distances between distributions via Stein's method. *J. Theoret. Probab.*, **35** (2), 949–987 (2022). MR4414410.

Fathi, M. Stein kernels and moment maps. *Ann. Probab.*, **47** (4), 2172–2185 (2019). MR3980918.

Fathi, M. Higher-order Stein kernels for Gaussian approximation. *Studia Math.*, **256** (3), 241–258 (2021). MR4172696.

Hillion, E., Johnson, O., and Yu, Y. A natural derivative on $[0, n]$ and a binomial Poincaré inequality. *ESAIM Probab. Stat.*, **18**, 703–712 (2014). MR3334010.

Houdré, C. and Kagan, A. Variance inequalities for functions of Gaussian variables. *J. Theoret. Probab.*, **8** (1), 23–30 (1995). MR1308667.

Houdré, C. and Pérez-Abreu, V. Covariance identities and inequalities for functionals on Wiener and Poisson spaces. *Ann. Probab.*, **23** (1), 400–419 (1995). MR1330776.

Houdré, C., Pérez-Abreu, V., and Surgailis, D. Interpolation, correlation identities, and inequalities for infinitely divisible variables. *J. Fourier Anal. Appl.*, **4** (6), 651–668 (1998). MR1665993.

Johnson, R. W. A note on variance bounds for a function of a Pearson variate. *Statist. Decisions*, **11** (3), 273–278 (1993). MR1257861.

Klaassen, C. A. J. On an inequality of Chernoff. *Ann. Probab.*, **13** (3), 966–974 (1985). MR799431.

Koldobsky, A. L. and Montgomery-Smith, S. J. Inequalities of correlation type for symmetric stable random vectors. *Statist. Probab. Lett.*, **28** (1), 91–97 (1996). MR1394423.

Korwar, R. M. On characterizations of distributions by mean absolute deviation and variance bounds. *Ann. Inst. Statist. Math.*, **43** (2), 287–295 (1991). MR1128869.

Ledoux, M. L'algèbre de Lie des gradients itérés d'un générateur markovien—développements de moyennes et entropies. *Ann. Sci. École Norm. Sup. (4)*, **28** (4), 435–460 (1995). MR1334608.

Ledoux, M., Nourdin, I., and Peccati, G. Stein's method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.*, **25** (1), 256–306 (2015). MR3320893.

Ley, C., Reinert, G., and Swan, Y. Stein's method for comparison of univariate distributions. *Probab. Surv.*, **14**, 1–52 (2017). MR3595350.

López Blázquez, F. and Salamanca Miño, B. Maximal correlation in a non-diagonal case. *J. Multivariate Anal.*, **131**, 265–278 (2014). MR3252649.

Miclo, L. Quand est-ce que des bornes de Hardy permettent de calculer une constante de Poincaré exacte sur la droite? *Ann. Fac. Sci. Toulouse Math. (6)*, **17** (1), 121–192 (2008). MR2464097.

Nash, J. Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.*, **80**, 931–954 (1958). MR100158.

Olkin, I. and Shepp, L. A matrix variance inequality. *J. Statist. Plann. Inference*, **130** (1-2), 351–358 (2005). MR2128013.

Papathanasiou, V. Variance bounds by a generalization of the Cauchy-Schwarz inequality. *Statist. Probab. Lett.*, **7** (1), 29–33 (1988). MR996849.

Pike, J. and Ren, H. Stein's method and the Laplace distribution. *ALEA Lat. Am. J. Probab. Math. Stat.*, **11** (1), 571–587 (2014). MR3283586.

Prakasa Rao, B. L. S. Matrix variance inequalities for multivariate distributions. *Stat. Methodol.*, **3** (4), 416–430 (2006). MR2252395.

Saumard, A. Weighted Poincaré inequalities, concentration inequalities and tail bounds related to Stein kernels in dimension one. *Bernoulli*, **25** (4B), 3978–4006 (2019). MR4010979.

Stein, C. *Approximate computation of expectations*, volume 7 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA (1986). ISBN 0-940600-08-0. MR882007.

Tanguy, K. *Quelques inégalités de superconcentration: théorie et applications*. Ph.D. thesis, Université Paul Sabatier-Toulouse III (2017).

Wei, Z. and Zhang, X. Covariance matrix inequalities for functions of Beta random variables. *Statist. Probab. Lett.*, **79** (7), 873–879 (2009). MR2509476.