Bernoulli **17**(2), 2011, 507–529
DOI: 10.3150/10-BEJ282

An extended Stein-type covariance identity for the Pearson family with applications to lower variance bounds

G. AFENDRAS[1], N. PAPADATOS[2,*] and V. PAPATHANASIOU[2,**]

[1]*Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus. E-mail: afendras.giorgos@ucy.ac.cy*
[2]*Department of Mathematics, Section of Statistics and O.R., University of Athens, Panepistemiopolis, 157 84 Athens, Greece. E-mail: [*]npapadat@math.uoa.gr; [**]bpapath@math.uoa.gr*

For an absolutely continuous (integer-valued) r.v. X of the Pearson (Ord) family, we show that, under natural moment conditions, a Stein-type covariance identity of order k holds (cf. [Goldstein and Reinert, *J. Theoret. Probab.* **18** (2005) 237–260]). This identity is closely related to the corresponding sequence of orthogonal polynomials, obtained by a Rodrigues-type formula, and provides convenient expressions for the Fourier coefficients of an arbitrary function. Application of the covariance identity yields some novel expressions for the corresponding lower variance bounds for a function of the r.v. X, expressions that seem to be known only in particular cases (for the Normal, see [Houdré and Kagan, *J. Theoret. Probab.* **8** (1995) 23–30]; see also [Houdré and Pérez-Abreu, *Ann. Probab.* **23** (1995) 400–419] for corresponding results related to the Wiener and Poisson processes). Some applications are also given.

Keywords: completeness; differences and derivatives of higher order; Fourier coefficients; orthogonal polynomials; Parseval identity; "Rodrigues inversion" formula; Rodrigues-type formula; Stein-type identity; variance bounds

1. Introduction

For an r.v. X with density f, mean μ and finite variance σ^2, Goldstein and Reinert [18] showed the identity (see also [26])

$$\operatorname{Cov}(X, g(X)) = \sigma^2 \mathbb{E}[g'(X^*)], \tag{1.1}$$

which holds for any absolutely continuous function $g : \mathbb{R} \to \mathbb{R}$ with a.s. derivative g' such that the right-hand side is finite. In (1.1), X^* is defined to be the r.v. with density $f^*(x) = \frac{1}{\sigma^2} \int_{-\infty}^{x} (\mu - t) f(t)\,\mathrm{d}t = \frac{1}{\sigma^2} \int_{x}^{\infty} (t - \mu) f(t)\,\mathrm{d}t$, $x \in \mathbb{R}$.

Identity (1.1) extends the well-known Stein identity for the standard normal [33,34]; a discrete version of (1.1) can be found in, for example, [13], where the derivative has been replaced by the forward difference of g. In particular, identities of the form (1.1) have many applications to variance bounds and characterizations [4,13,26], and to approximation procedures [11,14,15,18,27,31,33]. Several extensions and applications can be found in [10,19,29].

In [23], the (continuous) Pearson family is parametrized by the fact that there exists a quadratic $q(x) = \delta x^2 + \beta x + \gamma$ such that

$$\int_{-\infty}^{x} (\mu - t) f(t)\,\mathrm{d}t = q(x) f(x), \qquad x \in \mathbb{R}. \tag{1.2}$$

Typically, the usual definition of a Pearson r.v. is related to the differential equation $f'(x)/f(x) = (\alpha - x)/p_2(x)$, with p_2 being a polynomial of degree at most 2. In fact, the set-up of (1.2) (including, e.g., the standard uniform distribution with $q(x) = x(1-x)/2$) will be the framework of the present work and will hereafter be called "the Pearson family of continuous distributions". It is easily seen that under (1.2), the support of X, $S(X) = \{x : f(x) > 0\}$, must be an interval, say (r, s) with $-\infty \le r < s \le \infty$, and $q(x)$ remains strictly positive for $x \in (r, s)$. Clearly, under (1.2), the covariance identity (1.1) can be rewritten as

$$\mathbb{E}[(X - \mu)g(X)] = \mathbb{E}[q(X)g'(X)]. \tag{1.3}$$

It is known that, under appropriate moment conditions, the functions

$$P_k(x) = \frac{(-1)^k}{f(x)} \frac{\mathrm{d}^k}{\mathrm{d}x^k}[q^k(x) f(x)], \qquad x \in (r, s),\ k = 0, 1, \ldots, M \tag{1.4}$$

(where M can be finite or infinite) are orthogonal polynomials with respect to the density f so that the quadratic $q(x)$ in (1.2) generates a sequence of orthogonal polynomials by the Rodrigues-type formula (1.4). In fact, this approach is related to the Sturm–Liouville theory, [17], Section 5.2; see also [24,28].

In the present paper, we provide an extended Stein-type identity of order k for the Pearson family. This identity takes the form

$$\mathbb{E}[P_k(X)g(X)] = \mathbb{E}\big[q^k(X)g^{(k)}(X)\big], \tag{1.5}$$

where $g^{(k)}$ is the kth derivative of g (since $P_1(x) = x - \mu$, (1.5) for $k = 1$ reduces to (1.3)). Identity (1.5) provides a convenient formula for the kth Fourier coefficient of g, corresponding to the orthogonal polynomial P_k in (1.4). For its proof, we make use of a novel "Rodrigues inversion" formula that may be of some interest in itself. An identity similar to (1.5) holds for the discrete Pearson (Ord) family. Application of (1.5) and its discrete analog yields the corresponding lower variance bounds, obtained in Section 4. The lower bound for the Poisson(λ) distribution, namely,

$$\operatorname{Var} g(X) \ge \sum_{k=1}^{n} \frac{\lambda^k}{k!} \mathbb{E}^2[\Delta^k[g(X)]] \tag{1.6}$$

(cf. [21]) and the corresponding one for the Normal(μ, σ^2) distribution [20],

$$\operatorname{Var} g(X) \ge \sum_{k=1}^{n} \frac{(\sigma^2)^k}{k!} \mathbb{E}^2\big[g^{(k)}(X)\big], \tag{1.7}$$

are particular examples (Examples 4.1 and 4.5) of Theorems 4.1 and 4.2, respectively. Both (1.6) and (1.7) are particular cases of the finite form of Bessel's inequality and, under completeness, they can be extended to the corresponding Parseval identity. In Section 5, we show that this can be done for a fairly large family of r.v.'s, including, of course, the normal, the Poisson and, in general, all the r.v.'s of the Pearson system which have finite moments of any order. For instance, when X is Normal(μ, σ^2), inequality (1.7) (and identity (1.5)) can be strengthened to the covariance identity

$$\operatorname{Cov}[g_1(X), g_2(X)] = \sum_{k=1}^{\infty} \frac{(\sigma^2)^k}{k!} \mathbb{E}\big[g_1^{(k)}(X)\big]\mathbb{E}\big[g_2^{(k)}(X)\big], \tag{1.8}$$

provided that for $i = 1, 2$, $g_i \in \mathbb{D}^\infty(\mathbb{R})$, $\mathbb{E}|g_i^{(k)}(X)| < \infty$, $k = 1, 2, \ldots$, and that $\mathbb{E}[g_i(X)]^2 < \infty$. Similar identities hold for Poisson, negative binomial, beta and gamma distributions. These kinds of variance/covariance expressions may sometimes be useful in inference problems – see, e.g., the Applications 5.1 and 5.2 at the end of the paper.

2. Discrete orthogonal polynomials and the covariance identity

In order to simplify notation, we assume that X is a non-negative integer-valued r.v. with mean $\mu < \infty$. We also assume that there exists a quadratic $q(x) = \delta x^2 + \beta x + \gamma$ such that

$$\sum_{j=0}^{x} (\mu - j)p(j) = q(x)p(x), \qquad x = 0, 1, \ldots, \tag{2.1}$$

where $p(x)$ is the probability function of X. Relation (2.1) describes the discrete Pearson system (Ord family) [23]. Let Δ^k be the forward difference operator defined by $\Delta[g(x)] = g(x+1) - g(x)$ and $\Delta^k[g(x)] = \Delta[\Delta^{k-1}[g(x)]]$ ($\Delta^0[g] \equiv g$, $\Delta^1 \equiv \Delta$). We also set $q^{[k]}(x) = q(x)q(x+1)\cdots q(x+k-1)$ (with $q^{[0]} \equiv 1$, $q^{[1]} \equiv q$).

We first show some useful lemmas.

Lemma 2.1. *If $h(x) = 0$ for $x < 0$ and*

$$\sum_{x=0}^{\infty} |\Delta^j[h(x-j)]\Delta^{k-j}[g(x)]| < \infty \qquad \text{for } j = 0, 1, \ldots, k, \tag{2.2}$$

$$\lim_{x\to\infty} \Delta^j[h(x-j)]\Delta^{k-j-1}[g(x)] = 0 \qquad \text{for } j = 0, 1, \ldots, k-1, \tag{2.3}$$

then

$$\begin{aligned}(-1)^k &\sum_{x=0}^{\infty} \Delta^k[h(x-k)]g(x) \\ &= \sum_{x=0}^{\infty} h(x)\Delta^k[g(x)].\end{aligned} \tag{2.4}$$

Proof. We have

$$
\begin{aligned}
\sum_{x=0}^{\infty} h(x)\Delta^k[g(x)] &= \lim_{n\to\infty} \sum_{x=0}^{n} h(x)\big(\Delta^{k-1}[g(x+1)] - \Delta^{k-1}[g(x)]\big) \\
&= \lim_{n\to\infty}\left[h(n+1)\Delta^{k-1}[g(n+1)] - \sum_{x=0}^{n+1} \Delta[h(x-1)]\Delta^{k-1}[g(x)]\right] \\
&= \lim_{n\to\infty} h(n+1)\Delta^{k-1}[g(n+1)] - \sum_{x=0}^{\infty} \Delta[h(x-1)]\Delta^{k-1}[g(x)] \\
&= -\sum_{x=0}^{\infty} \Delta[h(x-1)]\Delta^{k-1}[g(x)].
\end{aligned}
$$

By the same calculation, it follows that

$$
(-1)^j \sum_{x=0}^{\infty} \Delta^j[h(x-j)]\Delta^{k-j}[g(x)] = (-1)^{j+1} \sum_{x=0}^{\infty} \Delta^{j+1}[h(x-j-1)]\Delta^{k-j-1}[g(x)]
$$

for any $j \in \{0, 1, \ldots, k-1\}$. □

Lemma 2.2. *For each $n \geq 0$, there exist polynomials $Q_{i,n}(x)$, $i = 0, 1, \ldots, n$, such that the degree of each $Q_{i,n}$ is at most i and*

$$
\Delta^i\big[q^{[n]}(x-n)p(x-n)\big] = q^{[n-i]}(x-n+i)p(x-n+i)Q_{i,n}(x), \qquad i = 0, 1, \ldots, n. \tag{2.5}
$$

Moreover, the leading coefficient (i.e., the coefficient of x^n) of $Q_{n,n}$ is given by $\mathrm{lead}(Q_{n,n}) = (-1)^n \prod_{j=n-1}^{2n-2}(1 - j\delta)$, *where an empty product should be treated as* 1.

Proof. For $n = 0$, the assertion is obvious and $Q_{0,0}(x) = 1$. For $n = 1$, the assertion follows from the assumption (2.1) with $Q_{0,1}(x) = 1$, $Q_{1,1}(x) = \mu - x$. For the case $n \geq 2$, the assertion will be proven using (finite) induction on i. Indeed, for $i = 0$, (2.5) holds with $Q_{0,n}(x) = 1$. Assuming that the assertion holds for some $i \in \{0, 1, \ldots, n-1\}$ and setting $h_n(x) = q^{[n]}(x-n)p(x-n)$, it follows that

$$
\begin{aligned}
\Delta^{i+1}h_n(x) &= \Delta[\Delta^i h_n(x)] \\
&= \Delta[h_{n-i}(x)Q_{i,n}(x)] \\
&= \Delta\big[q(x-n+i)p(x-n+i)\big(q^{[n-i-1]}(x-n+i+1)Q_{i,n}(x)\big)\big] \\
&= q(x-n+i+1)p(x-n+i+1)\Delta\big[q^{[n-i-1]}(x-n+i+1)Q_{i,n}(x)\big] \\
&\quad + \big(\mu - (x-n+i+1)\big)p(x-n+i+1)q^{[n-i-1]}(x-n+i+1)Q_{i,n}(x),
\end{aligned}
$$

where the obvious relation $\Delta[q(x)p(x)] = (\mu - (x+1))p(x+1)$, equivalent to (2.1), has been used with $x - n + i$ in place of x. Moreover,

$$\begin{aligned}\Delta&\big[q^{[n-i-1]}(x-n+i+1)Q_{i,n}(x)\big]\\ &= q^{[n-i-1]}(x-n+i+2)\Delta[Q_{i,n}(x)] + Q_{i,n}(x)\Delta\big[q^{[n-i-1]}(x-n+i+1)\big]\end{aligned}$$

and

$$\Delta\big[q^{[n-i-1]}(x-n+i+1)\big] = q^{[n-i-2]}(x-n+i+2)\big(q(x) - q(x-n+i+1)\big)$$

(where, for $i = n-1$, the right-hand side of the above should be treated as 0). Observing that

$$\begin{aligned}q(x-n+i+1)&q^{[n-i-2]}(x-n+i+2)\big(q(x)-q(x-n+i+1)\big)\\ &= q^{[n-i-1]}(x-n+i+1)\big(q(x)-q(x-n+i+1)\big)\end{aligned}$$

(which is also true in the case where $i = n-1$), the above calculations show that (2.5) holds with

$$Q_{i+1,n}(x) = P_{i,n}(x)Q_{i,n}(x) + q(x)\Delta[Q_{i,n}(x)],$$

where $P_{i,n}(x) = \mu - (x-n+i+1) + q(x) - q(x-n+i+1)$ is a linear polynomial or a constant. From the above recurrence, it follows immediately that $\mathrm{lead}(Q_{i+1:n}) = -(1-(2n-i-2)\delta)\,\mathrm{lead}(Q_{i,n})$, $i = 0, 1, \dots, n-1$; this, combined with the fact that $\mathrm{lead}(Q_{0,n}) = 1$, yields the desired result. □

It is easy to see that under (2.1), the support of X, $S(X) = \{x \in \mathbb{Z} : p(x) > 0\}$, is a finite or infinite integer interval. This integer interval will be denoted by J. Here, the term "integer interval" means that "if j_1 and j_2 are integers belonging to J, then all integers between j_1 and j_2 also belong to J".

Lemma 2.3. *For each $k = 0, 1, 2, \dots$, define the functions $P_k(x)$, $x \in J$, by the Rodrigues-type formula*

$$\begin{aligned}P_k(x) &= \frac{(-1)^k}{p(x)}\Delta^k\big[q^{[k]}(x-k)p(x-k)\big]\\ &= \frac{1}{p(x)}\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}q^{[k]}(x-j)p(x-j).\end{aligned} \tag{2.6}$$

We then have the following:

(a) *Each P_k is a polynomial of degree at most k, with* $\mathrm{lead}(P_k) = \prod_{j=k-1}^{2k-2}(1-j\delta)$ (in the sense that the function $P_k(x)$, $x \in J$, is the restriction of a real polynomial $G_k(x) = \sum_{j=0}^{k} c(k,j)x^j$, $x \in \mathbb{R}$, of degree at most k, such that $c(k,k) = \mathrm{lead}(P_k)$).

(b) *Let g be an arbitrary function defined on J* (the integer interval support of X) *and, if J is infinite, assume in addition that the functions g and $h(x) := q^{[k]}(x)p(x)$ satisfy the requirements* (2.2) *and* (2.3) *of Lemma* 2.1. *Then,*

$$\mathbb{E}|P_k(X)g(X)| < \infty, \qquad \mathbb{E}\big[q^{[k]}(X)|\Delta^k[g(X)]|\big] < \infty$$

and the following identity holds:

$$\mathbb{E}[P_k(X)g(X)] = \mathbb{E}\big[q^{[k]}(X)\Delta^k[g(X)]\big]. \tag{2.7}$$

Proof. Part (a) follows from Lemma 2.2 since $P_k = (-1)^k Q_{k,k}$. For part (b), assume first that $p(0) > 0$, that is, the support J is either $J = \{0, 1, \ldots\}$ or $J = \{0, 1, \ldots, N\}$ for a positive natural number N. In the unbounded case, (2.7) follows by an application of Lemma 2.1 to the functions $h(x) = q^{[k]}(x)p(x)$ and g since they satisfy the conditions (2.2) and (2.3). For the bounded case, it follows by (2.1) that $x = N$ is a zero of $q(x)$ so that $q(x) = (N - x)(\mu/N - \delta x)$, where, necessarily, $\delta < \mu/(N(N-1))$. Thus, in the case where $k > N$, $q^{[k]}(x) = 0$ for all $x \in J$ so that the right-hand side of (2.7) vanishes. On the other hand, the left-hand side of (2.7) (with P_k given by (2.6)) is $\sum_{x=0}^{N} g(x) \sum_{j=0}^{k} (-1)^{k-j}\binom{k}{j} q^{[k]}(x-j)p(x-j)$ and it is easy to verify that for all $x \in J$ and all $j \in \{0, 1, \ldots, k\}$, the quantity $q^{[k]}(x-j)p(x-j)$ vanishes. Thus, when $k > N$, (2.7) holds in the trivial sense $0 = 0$. For $k \le N$, the left-hand side of (2.7) equals

$$\begin{aligned}
&\sum_{j=0}^{k} (-1)^{k-j}\binom{k}{j} \sum_{x=-j}^{N-j} q^{[k]}(x)p(x)g(x+j) \\
&\quad = \sum_{j=0}^{k} (-1)^{k-j}\binom{k}{j} \sum_{x=0}^{N-j} q^{[k]}(x)p(x)g(x+j) \\
&\quad = \sum_{x=0}^{N} q^{[k]}(x)p(x) \sum_{j=0}^{\min\{k,N-x\}} (-1)^{k-j}\binom{k}{j} g(x+j) \\
&\quad = \sum_{x=0}^{N-k} q^{[k]}(x)p(x) \sum_{j=0}^{k} (-1)^{k-j}\binom{k}{j} g(x+j),
\end{aligned}$$

where we have made use of the facts that $p(x) = 0$ for $x < 0$ and $q^{[k]}(x) = 0$ for $N - k < x \le N$. Thus, (2.7) is proved under the assumption $p(0) > 0$. For the general case, where the support of X is either $\{m, m+1, \ldots, N\}$ or $\{m, m+1, \ldots\}$, it suffices to apply the same arguments to the r.v. $X - m$. $\square$

Theorem 2.1. *Suppose that X satisfies* (2.1) *and has $2n$ finite moments for some $n \ge 1$. The polynomials P_k, $k = 0, 1, \ldots, n$, defined by* (2.6), *then satisfy the orthogonality condition*

$$\mathbb{E}[P_k(X)P_m(X)] = \delta_{k,m} k! \mathbb{E}\big[q^{[k]}(X)\big] \prod_{j=k-1}^{2k-2} (1 - j\delta), \qquad k, m = 0, 1, \ldots, n, \tag{2.8}$$

where $\delta_{k,m}$ is Kronecker's delta.

Proof. Let $0 \le m \le k \le n$. First, observe that $q^{[k]}$ is a polynomial of degree at most $2k$ and P_m a polynomial of degree at most m. Note that the desired result will be deduced if we can apply (2.7) to the function $g = P_m$; this can be trivially applied when J is finite. In the remaining case where J is infinite, we have to verify conditions (2.2) and (2.3) of Lemma 2.1 for the functions $h = q^{[k]}p$ and $g = P_m$, or, more generally, for any polynomial g of degree less than or equal to k. Since the case $m = 0$ is obvious ($P_0(x) = 1$), we assume that $1 \le m \le k \le n$. Observe that (2.2) is satisfied in this case because of the assumption $\mathbb{E}|X|^{2k} < \infty$. Indeed, from Lemma 2.2, $\Delta^j[q^{[k]}(x-j)p(x-j)] = p(x)q^{[k-j]}(x)Q_{j,k}(x-j+k)$, where $Q_{j,k}$ is of degree at most j and $q^{[k-j]}$ is of degree at most $2k-2j$ so that their product is a polynomial of degree at most $2k-j$, while $\Delta^{k-j}[g]$ is a polynomial of degree at most j. On the other hand, since $q(x)p(x)$ is eventually decreasing and $\sum_{x=0}^{\infty} x^{2k-2}q(x)p(x) < \infty$, we have, as $y \to \infty$, that

$$q(2y)p(2y)y(y+1)^{2k-2} \le \sum_{x=y+1}^{2y} x^{2k-2}q(x)p(x) \to 0 \quad \text{and}$$

$$q(2y+1)p(2y+1)(y+1)^{2k-1} \le \sum_{x=y+1}^{2y+1} x^{2k-2}q(x)p(x) \to 0.$$

Hence, $\lim_{x\to\infty} x^{2k-1}q(x)p(x) = 0$. For $j \le k-1$, we have, from Lemma 2.2, that

$$\Delta^j[h(x-j)]\Delta^{k-j-1}[g(x)] = q(x)p(x)R(x),$$

where

$$R(x) = q^{[k-j-1]}(x+1)Q_{j,k}(x-j+k)\Delta^{k-j-1}[g(x)]$$

is a polynomial of degree at most $2k-1$. This, combined with the above limit, verifies (2.3) and a final application of (2.7) completes the proof. □

The covariance identity (2.7) enables the calculation of the Fourier coefficients of any function g in terms of its differences $\Delta^k[g]$, provided that conditions (2.2) and (2.3) are fulfilled for g and $h(x) = q^{[k]}(x)p(x)$. Since this identity is important for the applications, we state and prove a more general result that relaxes these conditions. The proof of this result contains a novel inversion formula for the polynomials obtained by the Rodrigues-type formula (2.6).

Theorem 2.2. *Assume that X satisfies (2.1) and has $2k$ finite moments, and suppose that for some function g defined on J,*

$$\mathbb{E}\big[q^{[k]}(X)|\Delta^k[g(X)]|\big] < \infty.$$

Then,

$$\mathbb{E}|P_k(X)g(X)| < \infty$$

and the covariance identity (2.7) *holds. Moreover, provided that* $k \ge 1$ *and* X *has* $2k-1$ *finite moments, the following "Rodrigues inversion" formula holds*:

$$q^{[k]}(x)p(x) = \frac{1}{(k-1)!} \sum_{y=x+1}^{\infty} (y-x-1)_{k-1} P_k(y) p(y), \qquad x = 0, 1, \ldots, \tag{2.9}$$

where, for $x \in \mathbb{R}$, $(x)_n = x(x-1)\cdots(x-n+1)$, $n = 1, 2, \ldots,$ *and* $(x)_0 = 1$.

Proof. Relation (2.7) is obvious when $k = 0$. Also, the assertion follows trivially by Lemma 2.3(b) if J is finite. Assume, now, that $k \ge 1$ and that J is unbounded, fix $s \in \{0, 1, \ldots\}$ and consider the function

$$g_{s,k}(x) := \frac{1}{(k-1)!} I(x \ge s+1)(x-s-1)_{k-1}, \qquad x = 0, 1, \ldots.$$

It is easily seen that the pair of functions $g = g_{s,k}$ and $h(x) = q^{[k]}(x)p(x)$ satisfy (2.2) and (2.3) and also that $\Delta^k[g_{s,k}(x)] = I(x = s)$. Thus, by (2.7), we get (2.9) (cf. formulae (3.6) and (3.7), below, for the continuous case), where P_k is defined by (2.6) and the series converges from Lemma 2.3(a) and the fact that X has $2k-1$ finite moments (note that $2k-1$ finite moments suffice for this inversion formula). Since P_k is a polynomial and the left-hand side of (2.9) is strictly positive for all large enough x, it follows that $P_k(y) > 0$ for all large enough y. Using (2.9), the assumption on $\Delta^k[g]$ is equivalent to the fact that

$$\frac{1}{(k-1)!} \sum_{x=0}^{\infty} |\Delta^k[g(x)]| \sum_{y=x+1}^{\infty} (y-x-1)_{k-1} P_k(y) p(y) < \infty$$

and arguments similar to those used in the proof of Theorem 3.1(b) below show that

$$\frac{1}{(k-1)!} \sum_{x=0}^{\infty} |\Delta^k[g(x)]| \sum_{y=x+1}^{\infty} (y-x-1)_{k-1} |P_k(y)| p(y) < \infty.$$

Therefore, we can interchange the order of summation, obtaining

$$\begin{aligned}
\mathbb{E}\bigl[q^{[k]}(X)\Delta^k[g(X)]\bigr] &= \frac{1}{(k-1)!} \sum_{x=0}^{\infty} \Delta^k[g(x)] \sum_{y=x+1}^{\infty} (y-x-1)_{k-1} P_k(y) p(y) \\
&= \frac{1}{(k-1)!} \sum_{y=1}^{\infty} P_k(y) p(y) \sum_{x=0}^{y-1} (y-x-1)_{k-1} \Delta^k[g(x)] \\
&= \mathbb{E}[P_k(X) G(X)],
\end{aligned}$$

where

$$G(x) = \frac{1}{(k-1)!} \sum_{y=0}^{x-k} (x-1-y)_{k-1} \Delta^k[g(y)], \qquad x = 0, 1, \ldots,$$

and where an empty sum should be treated as 0. Taking forward differences, it follows that $\Delta^k[G(x)] = \Delta^k[g(x)]$ so that $G = g + H_{k-1}$, where H_{k-1} is a polynomial of degree at most $k-1$, and the desired result follows from the orthogonality of P_k to polynomials of degree lower than k. This completes the proof. $\square$

3. The generalized Stein-type identity for the continuous case

The orthogonality of polynomials (1.4) has been shown, for example, in [17]; see also [24,28]. For our purposes, we review some details.

The induction formula, as in (2.5), here takes the form

$$\frac{\mathrm{d}^i}{\mathrm{d}x^i}[q^k(x)f(x)] = q^{k-i}(x)f(x)Q_{i,k}(x), \qquad i = 0, 1, \dots, k, \tag{3.1}$$

with $Q_{0,k}(x) = 1$, where $Q_{i,k}$ is a polynomial of degree at most i and the explicit recurrence for $Q_{i,k}$ is

$$Q_{i+1,k}(x) = \bigl(\mu - x + (k-i-1)(2\delta x + \beta)\bigr)Q_{i,k}(x) + q(x)Q'_{i,k}(x), \qquad i = 0, 1, \dots, k-1.$$

This immediately implies that $\operatorname{lead}(P_k) = (-1)^k \operatorname{lead}(Q_{k,k}) = \prod_{j=k-1}^{2k-2}(1 - j\delta)$, as in the discrete case; see also [7,17]. The covariance identity, as in (2.7), here takes the form (after repeated integration by parts; cf. [22,28])

$$\mathbb{E}[P_k(X)g(X)] = \mathbb{E}\bigl[q^k(X)g^{(k)}(X)\bigr], \tag{3.2}$$

provided that the expectations are finite and

$$\begin{aligned} &\lim_{x\to r+} q^{i+1}(x)f(x)Q_{k-i-1,k}(x)g^{(i)}(x) \\ &\quad = \lim_{x\to s-} q^{i+1}(x)f(x)Q_{k-i-1,k}(x)g^{(i)}(x) = 0, \qquad i = 0, 1, \dots, k-1 \end{aligned} \tag{3.3}$$

(here, $(r, s) = \{x : f(x) > 0\}$; that the support is an interval follows from (1.2)), where $Q_{i,k}$ are the polynomials defined by the recurrence above. Obviously, an alternative condition, sufficient for (3.3) (and, hence, also for (3.2)), is

$$\begin{aligned} &\lim_{x\to r+} q^{i+1}(x)f(x)x^j g^{(i)}(x) = \lim_{x\to s-} q^{i+1}(x)f(x)x^j g^{(i)}(x) = 0, \\ &\quad i = 0, 1, \dots, k-1,\ j = 0, 1, \dots, k-i-1. \end{aligned} \tag{3.4}$$

Assuming that X has $2k$ finite moments, $k \ge 1$, it is seen that (3.4), and hence (3.2), is fulfilled for any polynomial g of degree at most k. For example, for the upper limit, we have $\lim_{x\to s-} q(x)f(x) = 0$ so that (3.4) trivially holds if $s < \infty$; also, if $s = +\infty$, then (3.4) follows from $q(x)f(x) = \mathrm{o}(x^{-(2k-1)})$ as $x \to +\infty$, which can be shown by observing that

$q(x)f(x)$ is eventually decreasing, positive and, by the assumption of finite $2k$th moment, satisfies $\lim_{x\to+\infty}\int_{x/2}^{x} y^{2k-2}q(y)f(y)\,\mathrm{d}y = 0$. Therefore, the explicit orthogonality relation is

$$\mathbb{E}[P_k(X)P_m(X)] = \delta_{k,m}k!\mathbb{E}[q^k(X)]\prod_{j=k-1}^{2k-2}(1-j\delta), \qquad k,m = 0,1,\ldots,n, \tag{3.5}$$

where $\delta_{k,m}$ is Kronecker's delta, provided that X has $2n$ finite moments. The proof follows by a trivial application of (3.2) to $g(x) = P_m(x)$, for $0 \le m \le k \le n$ (cf. [28]).

It should be noted, however, that the condition (3.4) or (3.3) imposes some unnecessary restrictions on g. In fact, the covariance identity (3.2) (which enables a general form of the Fourier coefficients of g to be constructed in terms of its derivatives) holds, in our case, in its full generality; the proof requires the novel inversion formula (3.6) or (3.7) below, which may be of some interest in itself.

Theorem 3.1. *Assume that X satisfies* (1.2) *and consider the polynomial $P_k(x)$ defined by* (1.4), *where $(r,s) = \{x : f(x) > 0\}$.*

(a) *If X has $2k-1$ finite moments ($k \ge 1$), then the following "Rodrigues inversion" formula holds*:

$$q^k(x)f(x) = \frac{(-1)^k}{(k-1)!}\int_r^x (x-y)^{k-1}P_k(y)f(y)\,\mathrm{d}y \tag{3.6}$$

$$= \frac{1}{(k-1)!}\int_x^s (y-x)^{k-1}P_k(y)f(y)\,\mathrm{d}y, \qquad x \in (r,s). \tag{3.7}$$

(b) *If X has $2k$ finite moments and $\mathbb{E}q^k(X)|g^{(k)}(X)| < \infty$, then $\mathbb{E}|P_k(X)g(X)| < \infty$ and the covariance identity* (3.2) *holds.*

Proof. (a) Let $H_1(x)$, $H_2(x)$ be the left-hand side and right-hand side, respectively, of (3.6). It is easy to see that the integral $H_2(x)$ is finite (this requires only $2k-1$ finite moments). Moreover, expanding $(x-y)^{k-1}$ in the integrand of $H_2(x)$ according to Newton's formula, it follows that $H_2^{(k)}(x) = (-1)^k P_k(x)f(x)$, $x \in (r,s)$, and, thus, by the definition (1.4), $(H_1(x) - H_2(x))^{(k)} = H_1^{(k)}(x) - H_2^{(k)}(x)$ vanishes identically in (r,s). Therefore, $H_1 - H_2$ is a polynomial of degree at most $k-1$. By (3.1), $\lim_{x\to r+} H_1^{(i)}(x) = 0$ for all $i = 0,1,\ldots,k-1$ because $q(x)f(x) \to 0$ as $x \to r+$ and, for the case $r = -\infty$, $q(x)f(x) = \mathrm{o}(x^{-(2k-2)})$ as $x \to -\infty$ since the $(2k-1)$th moment is finite, q is of degree at most 2 and $q(x)f(x)$ is increasing and positive in a neighborhood of $-\infty$. Similarly, using the fact that P_k is a polynomial of degree at most k and observing that $\lim_{x\to r+} x^i\int_r^x y^{k+j-1-i}f(y)\,\mathrm{d}y = 0$ for all $i = 0,1,\ldots,k-1$, $j = 0,1,\ldots,k$ (again, $2k-1$ finite moments suffice for this conclusion), it follows that $\lim_{x\to r+} H_2^{(i)}(x) = 0$ for all $i = 0,1,\ldots,k-1$. This proves that $H_1 - H_2$ vanishes identically in (r,s) and (3.6) follows. Finally, (3.7) follows from (3.6) and (3.2) with $g(y) = (x-y)^{k-1}$ (the validity of (3.2) for polynomials g of degree at most $k-1$ can be shown directly, using repeated integration by parts, as above).

(b) Suppose that $k \geq 1$; otherwise, since $P_0(x) = 1$, we have nothing to show. Since $\mathbb{E}[P_k(X)] = \mathbb{E}[P_k(X)P_0(X)] = 0$ from (3.5), either $P_k(x)$ vanishes identically for $x \in (r,s)$ (in which case, (3.2) trivially holds) or, otherwise, it must change sign at least once in (r,s). Assume that P_k is not identically zero and consider the change-sign points of P_k in (r,s), say, $\rho_1 < \rho_2 < \cdots < \rho_m$ (of course, $1 \leq m \leq k$ because P_k is a polynomial of degree at most k). Fix a point ρ in the finite interval $[\rho_1, \rho_m]$ and write, with the help of (3.6) and (3.7),

$$\begin{aligned}
&\mathbb{E}\big[q^k(X)\big|g^{(k)}(X)\big|\big] \\
&\quad = \frac{(-1)^k}{(k-1)!}\int_r^{\rho} \big|g^{(k)}(x)\big| \int_r^x (x-y)^{k-1} P_k(y) f(y)\,\mathrm{d}y\,\mathrm{d}x \\
&\qquad + \frac{1}{(k-1)!}\int_{\rho}^{s} \big|g^{(k)}(x)\big| \int_x^s (y-x)^{k-1} P_k(y) f(y)\,\mathrm{d}y\,\mathrm{d}x.
\end{aligned} \tag{3.8}$$

Because of the assumption on g, both integrals on the right-hand side of (3.8) are finite. We wish to show that we can change the order of integration in both integrals in the right-hand side of (3.8). This will follow from Fubini's theorem if it can be shown that

$$I(\rho) = \int_{\rho}^{s} \big|g^{(k)}(x)\big| \int_x^s (y-x)^{k-1} |P_k(y)| f(y)\,\mathrm{d}y\,\mathrm{d}x < \infty \tag{3.9}$$

and similarly for the other integral in (3.8). Since $q(x)f(x) > 0$ for all $x \in (r,s)$ (see (1.2)), it follows that $q^k(x)f(x) > 0$ for all $x \in (r,s)$ as well. Thus, from (3.7), we get

$$\int_x^s (y-x)^{k-1} P_k(y) f(y)\,\mathrm{d}y > 0, \qquad x \in [\rho, s). \tag{3.10}$$

On the other hand, $P_k(x)$ does not change sign in the interval (ρ_m, s) and, hence, $P_k(x) > 0$ for all $x \in (\rho_m, s)$, showing that

$$\begin{aligned}
&\int_{\rho_m}^{s} \big|g^{(k)}(x)\big| \int_x^s (y-x)^{k-1} |P_k(y)| f(y)\,\mathrm{d}y\,\mathrm{d}x \\
&\quad = \int_{\rho_m}^{s} \big|g^{(k)}(x)\big| \int_x^s (y-x)^{k-1} P_k(y) f(y)\,\mathrm{d}y\,\mathrm{d}x.
\end{aligned} \tag{3.11}$$

By the above considerations, it follows that

$$H(\rho) := \inf_{x\in[\rho,\rho_m]} h(x) := \inf_{x\in[\rho,\rho_m]} \int_x^s (y-x)^{k-1} P_k(y) f(y)\,\mathrm{d}y > 0, \tag{3.12}$$

$$S(\rho) := \sup_{x\in[\rho,\rho_m]} s(x) := \sup_{x\in[\rho,\rho_m]} \int_x^{\rho_m} (y-x)^{k-1} |P_k(y)| f(y)\,\mathrm{d}y < \infty, \tag{3.13}$$

$$D(\rho) := \sup_{x\in[\rho,\rho_m]} d(x) := \sup_{x\in[\rho,\rho_m]} \int_{\rho_m}^{s} (y-x)^{k-1} P_k(y) f(y)\,\mathrm{d}y < \infty \tag{3.14}$$

because the three positive functions $h(x)$, $s(x)$ and $d(x)$ defined above are obviously continuous and x lies in the compact interval $[\rho, \rho_m]$ (note that $h(x) > 0$ by (3.10)). Now, from the inequalities $s(x) \le \frac{S(\rho)}{H(\rho)} h(x)$ and $d(x) \le \frac{D(\rho)}{H(\rho)} h(x)$, $x \in [\rho, \rho_m]$, we conclude that

$$\begin{aligned}
&\int_\rho^{\rho_m} \big|g^{(k)}(x)\big| \int_x^s (y-x)^{k-1} |P_k(y)| f(y)\,\mathrm{d}y\,\mathrm{d}x \\
&\quad = \int_\rho^{\rho_m} \big|g^{(k)}(x)\big| \big(s(x)+d(x)\big)\,\mathrm{d}x && (3.15)\\
&\quad \le \frac{S(\rho)+D(\rho)}{H(\rho)} \int_\rho^{\rho_m} \big|g^{(k)}(x)\big| \int_x^s (y-x)^{k-1} P_k(y) f(y)\,\mathrm{d}y\,\mathrm{d}x.
\end{aligned}$$

Combining (3.11) and (3.15), we see that there exists a finite constant $C(\rho)$ (take, for example, $C(\rho) = \max\{1, (S(\rho)+D(\rho))/H(\rho)\}$) such that

$$I(\rho) \le C(\rho) \int_\rho^s \big|g^{(k)}(x)\big| \int_x^s (y-x)^{k-1} P_k(y) f(y)\,\mathrm{d}y\,\mathrm{d}x < \infty$$

and, thus, by (3.9), we can indeed interchange the order of integration in the second integral in the right-hand side of (3.8). Similar arguments apply to the first integral. By the above arguments and by interchanging the order of integration in both integrals in the right-hand side of (3.8) (with $g^{(k)}$ in place of $|g^{(k)}|$), we obtain

$$\mathbb{E}\big[q^k(X) g^{(k)}(X)\big] = \mathbb{E}[P_k(X) G(X)], \tag{3.16}$$

where

$$\begin{aligned}
G(x) &= \frac{(-1)^k}{(k-1)!} \int_x^\rho (y-x)^{k-1} g^{(k)}(y)\,\mathrm{d}y && (3.17)\\
&= \frac{1}{(k-1)!} \int_\rho^x (x-y)^{k-1} g^{(k)}(y)\,\mathrm{d}y, \qquad x \in (r,s). && (3.18)
\end{aligned}$$

Differentiating (3.17) or (3.18) k times, it is easily seen that $G^{(k)} = g^{(k)}$ so that $G - g = H_{k-1}$ is a polynomial of degree at most $k-1$ and the desired result follows by (3.16) and the orthogonality of P_k to polynomials of degree lower than k. This completes the proof of the theorem. □

4. An application to lower variance bounds

A simple application of Theorem 2.2 leads to the following lower variance bound.

Theorem 4.1. *Fix $n \in \{1, 2, \dots\}$ and assume that X satisfies (2.1) and has $2n$ finite moments. Then, for any function g satisfying*

$$\mathbb{E}\big[q^{[k]}(X) |\Delta^k[g(X)]|\big] < \infty \qquad \textit{for } k = 0, 1, \dots, n, \tag{4.1}$$

the bound

$$\operatorname{Var} g(X) \geq \sum_{k=1}^{n} \frac{\mathbb{E}^2[q^{[k]}(X)\Delta^k[g(X)]]}{k!\mathbb{E}[q^{[k]}(X)]\prod_{j=k-1}^{2k-2}(1-j\delta)} \tag{4.2}$$

holds (where the kth term in the sum should be treated as zero whenever $\mathbb{E}[q^{[k]}(X)]$ vanishes) *with equality if and only if g is a polynomial of degree at most n.*

Proof. Assume that $\mathbb{E}[g^2(X)] < \infty$ (otherwise, we have nothing to show). By Theorem 2.1, the polynomials $\{P_k/\sqrt{\mathbb{E}[P_k^2(X)]}; k = 0, 1, \ldots, \min\{n, N\}\}$ form an orthonormal basis of all polynomials with degree up to n, where $N + 1$ is the cardinality of J. Observing that the kth Fourier coefficient for g is, by (2.7), (2.8) and Theorem 2.2,

$$\frac{\mathbb{E}[P_k(X)g(X)]}{\mathbb{E}^{1/2}[P_k^2(X)]} = \frac{\mathbb{E}[q^{[k]}(X)\Delta^k[g(X)]]}{(k!\mathbb{E}[q^{[k]}(X)]\prod_{j=k-1}^{2k-2}(1-j\delta))^{1/2}}, \qquad k \leq \min\{n, N\},$$

the desired result follows by an application of the finite form of Bessel's inequality. □

It is worth mentioning here the similarity of the lower variance bound (4.2) with the Poincaré-type (upper/lower) bound for the discrete Pearson family, obtained recently in [2], namely

$$(-1)^n \left(\operatorname{Var} g(X) - \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!\prod_{j=0}^{k-1}(1-j\delta)} \mathbb{E}\big[q^{[k]}(X)(\Delta^k[g(X)])^2\big] \right) \geq 0. \tag{4.3}$$

The following examples can be verified immediately.

Example 4.1. If X is Poisson(λ), then $q(x) = \lambda$ so that $\delta = 0$ and (1.6) follows from (4.2) (see also [21]). Moreover, the equality in (1.6) holds if and only if g is a polynomial of degree at most n.

Example 4.2. For the binomial(N, p) distribution, $q(x) = (N - x)p$ so that $\delta = 0$ and $\mathbb{E}[q^{[k]}(X)] = (N)_k p^k (1-p)^k$. Thus, (4.2) yields the bound

$$\operatorname{Var} g(X) \geq \sum_{k=1}^{\min\{n,N\}} \frac{p^k}{k!(N)_k(1-p)^k} \mathbb{E}^2\big[(N-X)_k \Delta^k[g(X)]\big]$$

with equality only for polynomials of degree at most n. Note that there is equality if $n \geq N$.

Example 4.3. For the negative binomial(r, p) with $p(x) = (r + x - 1)_x p^r (1-p)^x / x!$, $x = 0, 1, \ldots$, (2.1) is satisfied with $q(x) = (1-p)(r+x)/p$. Thus, $\delta = 0$, $\mathbb{E}[q^{[k]}(X)] = (1-p)^k[r]_k/p^{2k}$ and (4.2) produces the bound

$$\operatorname{Var} g(X) \geq \sum_{k=1}^{n} \frac{(1-p)^k}{k![r]_k} \mathbb{E}^2\big[[r+X]_k \Delta^k[g(X)]\big]$$

(in the above formulae and elsewhere in the paper, $[x]_n = x(x+1)\cdots(x+n-1)$ if $n \ge 1$ and $[x]_0 = 1$).

Example 4.4. For the discrete Uniform$\{1, 2, \dots, N\}$, (2.1) is satisfied with $q(x) = x(N-x)/2$; thus, $\delta = -1/2$ and (4.2) entails the bound (which is an identity if $n \ge N-1$)

$$\operatorname{Var} g(X) \ge N \sum_{k=1}^{\min\{n,N-1\}} \frac{(2k+1)(N-k-1)!}{(k!)^2(N+k)!} \mathbb{E}^2\big[[X]_k (N-X)_k \Delta^k [g(X)]\big].$$

The lower variance bound for the continuous Pearson system is stated in the following theorem; its proof, being an immediate consequence of (3.2), (3.5), Theorem 3.1 and a straightforward application of the finite form of Bessel's inequality (cf. the proof of Theorem 4.1 above), is omitted.

Theorem 4.2. *Assume that X satisfies* (1.2) *and has finite moment of order $2n$ for some fixed $n \ge 1$. Then, for any function g satisfying $\mathbb{E}[q^k(X)|g^{(k)}(X)|] < \infty$, $k = 0, 1, \dots, n$, we have the inequality*

$$\operatorname{Var} g(X) \ge \sum_{k=1}^{n} \frac{\mathbb{E}^2[q^k(X) g^{(k)}(X)]}{k!\mathbb{E}[q^k(X)] \prod_{j=k-1}^{2k-2}(1-j\delta)} \tag{4.4}$$

with equality if and only if g is a polynomial of degree at most n. (Note that $\mathbb{E}|X|^{2n} < \infty$ implies $\delta < (2n-1)^{-1}$ and, thus, that $\delta \notin \{1, 1/2, \dots, 1/(2n-2)\}$ if $n \ge 2$.)

Some examples now follow.

Example 4.5. If X is Normal(μ, σ^2), then $q(x) = \sigma^2$ and (4.4) yields the Houdré–Kagan variance bound (1.7) (see [20]), under the weaker assumptions $\mathbb{E}|g^{(k)}(X)| < \infty$, $k = 0, 1, \dots, n$; equality holds if and only if g is a polynomial of degree at most n.

Example 4.6. If X is $\Gamma(a, \lambda)$ with density $\lambda^a x^{a-1} \mathrm{e}^{-\lambda x}/\Gamma(a)$, $x > 0$, then $q(x) = x/\lambda$, $\mathbb{E}[q^k(X)] = [a]_k/\lambda^{2k}$ and (4.4) entails the bound

$$\operatorname{Var} g(X) \ge \sum_{k=1}^{n} \frac{1}{k![a]_k} \mathbb{E}^2\big[X^k g^{(k)}(X)\big]$$

with equality only for polynomials of degree at most n.

Example 4.7. For the standard uniform density, $q(x) = x(1-x)/2$, $\delta = -1/2$, $\mathbb{E}[q^k(X)] = (k!)^2/(2^k(2k+1)!)$ and we get the bound

$$\operatorname{Var} g(X) \ge \sum_{k=1}^{n} \frac{2k+1}{(k!)^2} \mathbb{E}^2\big[X^k (1-X)^k g^{(k)}(X)\big]$$

with equality only for polynomials of degree at most n. Similar inequalities hold for all beta densities.

Of course, in the above three examples, the corresponding orthogonal polynomials (1.4) are the well-known Hermite, Laguerre and Jacobi (Legendre), respectively, so that one can alternatively obtain the results using explicit expressions for the variance and generating functions for the polynomials (see, e.g., [5,6,12,16,19]); this is also the case for the discrete polynomials corresponding to Examples 4.1–4.4 (namely, Charlier, Krawtchouck, Meixner and Hahn, respectively). However, as will become clear from the following example, the considerations given here do not only simplify and unify the calculations, but also go beyond the classical polynomials.

Example 4.8. If X follows the t_N distribution (Student's t with N degrees of freedom) with density

$$f(x) = \frac{\Gamma((N+1)/2)}{\sqrt{N\pi}\,\Gamma(N/2)}\left(1+\frac{x^2}{N}\right)^{-(N+1)/2}, \qquad x \in \mathbb{R},$$

then it is well known that X has only $N-1$ finite integral moments. However, for $N > 1$, X satisfies (1.2) and its quadratic $q(x) = (N + x^2)/(N-1)$ has $\delta = 1/(N-1)$. Thus, (4.4) applies for sufficiently large N (see also [22] for a Poincaré-type bound corresponding to (4.3)). To this end, it suffices to calculate

$$\mathbb{E}[q^k(X)] = \left(\frac{N}{N-1}\right)^k \prod_{j=1}^{k}\left(1+\frac{1}{N-2j}\right), \qquad k \le (N-1)/2,$$

and $\prod_{j=k-1}^{2k-2}(1 - j\delta) = (N-k)_k/(N-1)^k$. Theorem 4.2 yields the (non-classic) bound

$$\operatorname{Var} g(X) \ge \sum_{k=1}^{n} \frac{\mathbb{E}^2[(N+X^2)^k g^{(k)}(X)]}{k!N^k(N-k)_k \prod_{j=1}^{k}(1+1/(N-2j))}, \qquad n \le \frac{N-1}{2},$$

with equality only for polynomials of degree at most n.

It seems that it would be difficult to work with the explicit forms of the corresponding orthogonal polynomials, obtained by (1.4). Note that a similar bound can be easily obtained for the Fisher–Snedecor F_{n_1,n_2} distribution and that Schoutens [31] has obtained the corresponding Stein's equation, useful in approximating the t_N-distribution.

5. A general variance/covariance representation

The main application of the present article (Section 4) presents a convenient procedure for approximating/bounding the variance of $g(X)$ when g is smooth enough and when X is "nice" enough (Pearson). Although the procedure is based on the corresponding orthonormal polynomials, $\phi_k = P_k/\mathbb{E}^{1/2}[P_k^2]$, the main point is that we do not need explicit forms for P_k. All we need are the Fourier coefficients of g, $c_k = \mathbb{E}[\phi_k g]$, but, due to the identities (3.2) and (2.7), the Fourier coefficients can be simply expressed in terms of the quadratic q and the derivatives/differences of g when X belongs to the Pearson/Ord family, that is, when (1.2) or (2.1) is satisfied.

A natural question thus arises: *is it true that the nth partial sums, given by the right-hand sides of* (4.2) *and* (4.4)*, converge to the variance of $g(X)$ as $n \to \infty$*? More generally, assume that we want to calculate $\operatorname{Var} g(X)$, when X is an r.v. with finite moments of any order. Let μ be the probability measure of X, that is, μ satisfies $\mu(-\infty, x] = \Pr(X \le x)$, $x \in \mathbb{R}$, and assume that the support of μ, $\operatorname{supp}(\mu)$, is not concentrated on a finite number of points (otherwise, the following considerations become trivial). It is well known that there exists an orthonormal polynomial system (OPS) $\mathcal{F} = \{\phi_0, \phi_1, \ldots\}$, which can be obtained by an application of the Gram–Schmidt orthonormalization process to the real system $\mathcal{F}_0 = \{1, x, x^2, \ldots\} \subset L^2(\mathbb{R}, \mu)$. Each (real) polynomial ϕ_k is of degree (exactly) k and satisfies the orthonormality condition $\mathbb{E}[\phi_i(X)\phi_j(X)] = \int_{\mathbb{R}} \phi_i(x)\phi_j(x)\,\mathrm{d}\mu(x) = \delta_{ij}$. The members of $\mathcal{F}$ are uniquely determined by the moments of X (of course, any element of $\mathcal{F}$ can be multiplied by ± 1). For any function $g \in L^2(\mathbb{R}, \mu)$ (i.e., with finite variance), we first calculate the Fourier coefficients

$$c_k = \mathbb{E}[g(X)\phi_k(X)] = \int_{\mathbb{R}} \phi_k(x) g(x)\,\mathrm{d}\mu(x), \qquad k = 0, 1, 2, \ldots, \tag{5.1}$$

and then use the well-known Bessel inequality,

$$\operatorname{Var} g(X) \ge \sum_{k=1}^{\infty} c_k^2, \tag{5.2}$$

to obtain the desirable lower variance bound. Clearly, Theorems 4.1 and 4.2 just provide convenient forms of the nth partial sum in (5.2) for some particularly interesting cases (Pearson/Ord system). It is well known that the sum in the right-hand side of (5.2) is equal to the variance (for all $g \in L^2(\mathbb{R}, \mu)$) if and only if the OPS $\mathcal{F}$ is complete in $L^2(\mathbb{R}, \mu)$ (and, thus, it is an orthonormal basis of $L^2(\mathbb{R}, \mu)$). This means that the set of real polynomials, $\operatorname{span}[\mathcal{F}_0]$, is dense in $L^2(\mathbb{R}, \mu)$. If this is the case, then Bessel's inequality, (5.2), is strengthened to Parseval's identity,

$$\operatorname{Var} g(X) = \sum_{k=1}^{\infty} c_k^2 \qquad \text{for any } g \in L^2(\mathbb{R}, \mu). \tag{5.3}$$

The following theorem summarizes and unifies the above observations.

Theorem 5.1. *Assume that X has probability measure μ and finite moments of any order.*

(a) *If X satisfies* (1.2) *and if $g \in L^2(\mathbb{R}, \mu) \cap D^\infty(r, s)$, then* (5.3) *can be written as* (cf. Theorem 4.2)

$$\operatorname{Var} g(X) = \sum_{k=1}^{\infty} \frac{\mathbb{E}^2[q^k(X) g^{(k)}(X)]}{k!\,\mathbb{E}[q^k(X)] \prod_{j=k-1}^{2k-2}(1 - j\delta)}, \tag{5.4}$$

provided that the polynomials are dense in $L^2(\mathbb{R}, \mu)$ and that

$$\mathbb{E}\big[q^k(X)\big|g^{(k)}(X)\big|\big] < \infty, \qquad k = 0, 1, 2, \ldots. \tag{5.5}$$

(b) *Similarly, if X satisfies* (2.1), *if μ is not concentrated on a finite integer interval and if $g \in L^2(\mathbb{R}, \mu)$, then* (5.3) *yields the identity* (cf. Theorem 4.1)

$$\operatorname{Var} g(X) = \sum_{k=1}^{\infty} \frac{\mathbb{E}^2[q^{[k]}(X)\Delta^k[g(X)]]}{k!\mathbb{E}[q^{[k]}(X)] \prod_{j=k-1}^{2k-2}(1 - j\delta)}, \tag{5.6}$$

provided that the polynomials are dense in $L^2(\mathbb{R}, \mu)$ and that

$$\mathbb{E}\big[q^{[k]}(X)|\Delta^k[g(X)]|\big] < \infty, \qquad k = 0, 1, 2, \ldots. \tag{5.7}$$

It should be noted that (5.6) is always true (and reduces to a finite sum) if X belongs to the discrete Pearson (Ord) family and the support of μ is finite. In this case, the sum adds zero terms whenever $\mathbb{E}[q^{[k]}(X)] = 0$. On the other hand, it is well known (due to M. Riesz) that the polynomials are dense in $L^2(\mathbb{R}, \mu)$ whenever μ is determined by its moments; see [30] or [3], page 45. An even simpler sufficient condition is when μ has a finite moment generating function at a neighborhood of zero, that is, when there exists $t_0 > 0$ such that

$$M_X(t) = \mathbb{E}\mathrm{e}^{tX} < \infty, \qquad t \in (-t_0, t_0). \tag{5.8}$$

A proof can be found in, for example, [8]. Since condition (5.8) can evidently be checked for all Pearson distributions, we include an alternative proof in the Appendix.

Taking into account the above, we have the following covariance representation.

Theorem 5.2. *Assume that X has a finite moment generating function at a neighborhood of zero.*

(a) *If X satisfies* (1.2) *and if $g_i \in L^2(\mathbb{R}, \mu) \cap D^\infty(r, s)$, $i = 1, 2$, then*

$$\operatorname{Cov}[g_1(X), g_2(X)] = \sum_{k=1}^{\infty} \frac{\mathbb{E}[q^k(X) g_1^{(k)}(X)] \cdot \mathbb{E}[q^k(X) g_2^{(k)}(X)]}{k!\mathbb{E}[q^k(X)] \prod_{j=k-1}^{2k-2}(1 - j\delta)}, \tag{5.9}$$

provided that for $i = 1, 2$,

$$\mathbb{E}\big[q^k(X)\big|g_i^{(k)}(X)\big|\big] < \infty, \qquad k = 0, 1, 2, \ldots. \tag{5.10}$$

(b) *Similarly, if X satisfies* (2.1), *then*

$$\operatorname{Cov}[g_1(X), g_2(X)] = \sum_{k=1}^{\infty} \frac{\mathbb{E}[q^{[k]}(X)\Delta^k[g_1(X)]] \cdot \mathbb{E}[q^{[k]}(X)\Delta^k[g_2(X)]]}{k!\mathbb{E}[q^{[k]}(X)] \prod_{j=k-1}^{2k-2}(1 - j\delta)}, \tag{5.11}$$

where each term with $\mathbb{E}[q^{[k]}(X)] = 0$ should be treated as zero, provided that for $i = 1, 2$,

$$\mathbb{E}\big[q^{[k]}(X)|\Delta^k[g_i(X)]|\big] < \infty, \qquad k = 0, 1, 2, \ldots. \tag{5.12}$$

Proof. Let $\alpha_k = \mathbb{E}[\phi_k(X)g_1(X)]$ and $\beta_k = \mathbb{E}[\phi_k(X)g_2(X)]$ be the Fourier coefficients of g_1 and g_2. It is then a standard inner product property in the Hilbert space that $\mathrm{Cov}[g_1(X), g_2(X)] = \sum_{k=1}^{\infty} \alpha_k\beta_k$. Substituting, for example,

$$\alpha_k = \mathbb{E}[\phi_k(X)g_1(X)] = \frac{\mathbb{E}[P_k(X)g_1(X)]}{\mathbb{E}^{1/2}[P_k^2(X)]} = \frac{\mathbb{E}[q^{[k]}(X)\Delta^k[g_1(X)]]}{(k!\mathbb{E}[q^{[k]}(X)\prod_{j=k-1}^{2k-2}(1-j\delta)])^{1/2}}$$

(and similarly for the continuous case and for β_k), we obtain (5.9) and (5.11). □

Application 5.1. Assume that $X_1, X_2, \ldots, X_\nu$ is a random sample from Geometric(θ), $0<\theta<1$, with probability function

$$p(x) = \theta(1-\theta)^x, \qquad x = 0, 1, \ldots,$$

and let $X = X_1 + \cdots + X_\nu$ be the complete sufficient statistic. The uniformly minimum variance unbiased estimator, UMVUE, of $-\log(\theta)$ is then given by (see [2])

$$T_\nu = T_\nu(X) = \begin{cases} 0, & \text{if } X = 0, \\ \dfrac{1}{\nu} + \dfrac{1}{\nu+1} + \cdots + \dfrac{1}{\nu+X-1}, & \text{if } X \in \{1, 2, \ldots\}. \end{cases}$$

Since no simple form exists for the variance of T_ν, the inequalities (4.3) have been used in [2] in order to prove asymptotic efficiency. However, X is negative binomial(ν, θ) and

$$\Delta^k[T_\nu(X)] = \frac{(-1)^{k-1}(k-1)!}{[\nu+X]_k}, \qquad k = 1, 2, \ldots,$$

so that one finds from (5.6) the exact expression (cf. Example 4.3)

$$\mathrm{Var}\, T_\nu = \sum_{k=1}^{\infty} \frac{(1-\theta)^k}{k^2\binom{\nu+k-1}{k}}.$$

Observe that the first term in the series, $(1-\theta)/\nu$, is the Cramér–Rao lower bound.

Now, let $W_{\nu;n} = W_{\nu;n}(X) = [\nu+X]_n/[\nu]_n$ be the UMVUE of θ^{-n} ($n = 1, 2, \ldots$) and $U_{\nu;n} = U_{\nu;n}(X) = (\nu-1)_n/[\nu-n+X]_n$ be the UMVUE of θ^n ($n = 1, 2, \ldots, \nu-1$). ($W_{\nu;n}(X)$ is a polynomial, of degree n, in X.) It follows that

$$\Delta^k[W_{\nu;n}(X)] = \begin{cases} \dfrac{(n)_k[\nu+X+k]_{n-k}}{[\nu]_n}, & k = 0, 1, \ldots, n, \\ 0, & k = n+1, n+2, \ldots, \end{cases}$$

$$\Delta^k[U_{\nu;n}(X)] = \frac{(-1)^k[n]_k(\nu-1)_n}{[\nu-n+X]_{n+k}}, \qquad k = 0, 1, 2, \ldots,$$

so that

$$\mathbb{E}\big[[\nu+X]_k \Delta^k[W_{\nu;n}(X)]\big] = (n)_k \mathbb{E}[W_{\nu;n}(X)] = (n)_k \theta^{-n}, \qquad k=0,1,\ldots,n,$$
$$\mathbb{E}\big[[\nu+X]_k \Delta^k[U_{\nu;n}(X)]\big] = (-1)^k [n]_k \mathbb{E}[U_{\nu;n}(X)] = (-1)^k [n]_k \theta^{n}, \qquad k=0,1,\ldots.$$

Using (2.7), (2.8), (5.6), (5.11) and Example 4.3, we immediately obtain the formulae

$$\mathrm{Cov}[T_\nu, W_{\nu;n}] = \theta^{-n} \sum_{k=1}^{n} (-1)^{k-1} \frac{(n)_k}{k[\nu]_k} (1-\theta)^k, \qquad n=1,2,\ldots,$$

$$\mathrm{Cov}[T_\nu, U_{\nu;n}] = -\theta^{n} \sum_{k=1}^{\infty} \frac{[n]_k}{k[\nu]_k} (1-\theta)^k, \qquad n=1,2,\ldots,\nu-1,$$

$$\mathrm{Cov}[W_{\nu;n}, W_{\nu;m}] = \theta^{-n-m} \sum_{k=1}^{\min\{n,m\}} \frac{(n)_k (m)_k}{k![\nu]_k} (1-\theta)^k, \qquad n,m=1,2,\ldots,$$

$$\mathrm{Cov}[U_{\nu;n}, U_{\nu;m}] = \theta^{n+m} \sum_{k=1}^{\infty} \frac{[n]_k [m]_k}{k![\nu]_k} (1-\theta)^k, \qquad n,m=1,2,\ldots,\nu-1,$$

$$\mathrm{Cov}[W_{\nu;n}, U_{\nu;m}] = \theta^{m-n} \sum_{k=1}^{n} (-1)^k \frac{(n)_k [m]_k}{k![\nu]_k} (1-\theta)^k,$$
$$n=1,2,\ldots,\ m=1,2,\ldots,\nu-1.$$

The above series expansions are in accordance with the corresponding results on Bhattacharyya bounds given in [9]; these results are also based on orthogonality and completeness properties of Bhattacharyya functions, obtained by Seth [32]. Similar series expansions for the variance can be found in [1,25].

Next, we present a similar application for the exponential distribution.

Application 5.2. Assume that $X_1, X_2, \ldots, X_\nu$ is a random sample from $\mathrm{Exp}(\lambda)$, $\lambda > 0$, with density $f(x) = \lambda \mathrm{e}^{-\lambda x}$, $x > 0$, and let $X = X_1 + \cdots + X_\nu$ be the complete sufficient statistic. We wish to obtain the UMVUE of $\log(\lambda)$ and its variance. Setting $U = \log(X_1)$, we find that

$$\mathbb{E}U = \int_0^\infty \mathrm{e}^{-x} \log(x)\,\mathrm{d}x - \log(\lambda) = -\gamma - \log(\lambda),$$

where $\gamma = 0.5772\ldots$ is Euler's constant. Therefore, $-\gamma - U$ is unbiased and it follows that the UMVUE of $\log(\lambda)$ is of the form

$$L_\nu = L_\nu(X) = \mathbb{E}[-\gamma - \log(X_1)|X] = -\log(X) - \gamma + \sum_{j=1}^{\nu-1} \frac{1}{j}.$$

Since X follows a $\Gamma(\nu,\lambda)$ distribution and $L_\nu^{(k)}(X)=(-1)^k(k-1)!X^{-k}$, we obtain from (5.4) (cf. Example 4.6) the formula

$$\operatorname{Var} L_\nu=\sum_{k=1}^{\infty}\frac{1}{k^2\binom{\nu+k-1}{k}},$$

which is quite similar to the formula for $\operatorname{Var} T_\nu$ in Application 5.1. (Once again, the first term in the series, $1/\nu$, is the Cramér–Rao bound.) Moreover, the series $\sum_{k=1}^{\infty}\left(k^2\binom{\nu+k-1}{k}\right)^{-1}$ can be simplified in a closed form. Indeed, observing that $\operatorname{Var} L_1=\sum_{k\ge 1}1/k^2=\pi^2/6$ and taking into account the identity

$$\frac{1}{k^2}\left(\binom{\nu+k-1}{k}^{-1}-\binom{\nu+k}{k}^{-1}\right)=\frac{(\nu-1)!}{\nu}\left(\frac{1}{[k]_\nu}-\frac{1}{[k+1]_\nu}\right),\qquad \nu,k=1,2,\ldots,$$

we have

$$\operatorname{Var} L_\nu-\operatorname{Var} L_{\nu+1}=\frac{(\nu-1)!}{\nu}\sum_{k=1}^{\infty}\left(\frac{1}{[k]_\nu}-\frac{1}{[k+1]_\nu}\right)=\frac{1}{\nu^2}$$

so that

$$\operatorname{Var} L_\nu=\frac{\pi^2}{6}-1-\frac{1}{2^2}-\cdots-\frac{1}{(\nu-1)^2}=\sum_{k\ge\nu}\frac{1}{k^2}.$$

Finally, using the last expression and the fact that $(k+1)^{-2}<\int_k^{k+1}x^{-2}\,\mathrm{d}x<k^{-2}$, we get the inequalities $\sum_{k\ge\nu}(k+1)^{-2}=\operatorname{Var} L_\nu-\nu^{-2}<\nu^{-1}<\operatorname{Var} L_\nu$, that is,

$$1<\nu\operatorname{Var} L_\nu<1+1/\nu,$$

which shows that L_ν is asymptotically efficient.

Appendix

A completeness proof under (5.8). It is well known that, under (5.8), X has finite moments of any order so that the OPS exists (and is unique). From the general theory of Hilbert spaces, it is known that $\mathcal{F}$ is a basis (i.e., it is complete) if and only if it is total, that is, if and only if there does not exist a non-zero function $g\in L^2(\mathbb{R},\mu)$ such that g is orthogonal to each ϕ_k. Therefore, it suffices to show that if $g\in L^2(\mathbb{R},\mu)$ and if

$$\mathbb{E}[g(X)\phi_k(X)]=0\qquad\text{for all } k=0,1,\ldots,\tag{A.1}$$

then $\Pr(g(X)=0)=1$. Since each ϕ_k is a polynomial with non-zero leading coefficient, (A.1) is equivalent to

$$\mathbb{E}[X^n g(X)] = \int_{\mathbb{R}} x^n g(x)\,\mathrm{d}\mu(x) = 0, \qquad n = 0, 1, \ldots, \tag{A.2}$$

and it thus suffices to prove that if (A.2) holds, then $\mathbb{E}|g(X)| = 0$. Since, by assumption, the functions $x \mapsto x^n$ and $x \mapsto \mathrm{e}^{tx}$ ($|t| < t_0/2$) belong to $L^2(\mathbb{R}, \mu)$, it follows from the Cauchy–Schwarz inequality that $\mathbb{E}[\mathrm{e}^{tX}|g(X)|] < \infty$ for $|t| < t_0/2$ and, thus, that

$$\mathbb{E}\big[\mathrm{e}^{|tX|}|g(X)|\big] \le \mathbb{E}[\mathrm{e}^{-tX}|g(X)|] + \mathbb{E}[\mathrm{e}^{tX}|g(X)|] < \infty, \qquad |t| < t_0/2. \tag{A.3}$$

From Beppo Levi's theorem and (A.3) it follows that

$$\sum_{n=0}^{\infty} \mathbb{E}\left[\frac{|tX|^n}{n!}|g(X)|\right] = \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{|tX|^n}{n!}|g(X)|\right] = \mathbb{E}\big[\mathrm{e}^{|tX|}|g(X)|\big] < \infty, \qquad |t| < t_0/2,$$

and, therefore, from Fubini's theorem and (A.2), we get

$$\mathbb{E}[\mathrm{e}^{tX} g(X)] = \mathbb{E}\left[\sum_{n=0}^{\infty} \frac{t^n X^n}{n!} g(X)\right] = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbb{E}[X^n g(X)] = 0, \qquad |t| < t_0/2. \tag{A.4}$$

Write $g^+(x) = \max\{g(x), 0\}$, $g^-(x) = \max\{-g(x), 0\}$ and observe that from (A.2) with $n = 0$, $\mathbb{E}g(X) = 0$. It follows that $\mathbb{E}g^+(X) = \mathbb{E}g^-(X) = \theta$, say, where $\theta = \mathbb{E}|g(X)|/2$. Clearly, $0 \le \theta < \infty$ because both g^+ and g^- are dominated by $|g|$ and, by assumption, $|g| \in L^2(\mathbb{R}, \mu) \subset L^1(\mathbb{R}, \mu)$. We assume now that $\theta > 0$ and we shall obtain a contradiction. Under $\theta > 0$, we can define two Borel probability measures, ν^+ and ν^-, as follows:

$$\nu^+(A) = \frac{1}{\theta}\int_A g^+(x)\,\mathrm{d}\mu(x), \qquad \nu^-(A) = \frac{1}{\theta}\int_A g^-(x)\,\mathrm{d}\mu(x), \qquad A \in \mathcal{B}(\mathbb{R}).$$

By definition, both ν^+ and ν^- are absolutely continuous with respect to μ, with Radon–Nikodym derivatives

$$\frac{\mathrm{d}\nu^+}{\mathrm{d}\mu} = \frac{1}{\theta}g^+(x), \qquad \frac{\mathrm{d}\nu^-}{\mathrm{d}\mu} = \frac{1}{\theta}g^-(x), \qquad x \in \mathbb{R}.$$

Since, by (A.4),

$$\int_{\mathbb{R}} \mathrm{e}^{tx} g^+(x)\,\mathrm{d}\mu(x) = \int_{\mathbb{R}} \mathrm{e}^{tx} g^-(x)\,\mathrm{d}\mu(x), \qquad |t| < t_0/2, \tag{A.5}$$

it follows that the moment generating functions of ν^+ and ν^- are finite (for $|t| < t_0/2$) and identical because, from (A.5), we have that, for any $t \in (-t_0/2, t_0/2)$,

$$\int_{\mathbb{R}} \mathrm{e}^{tx}\,\mathrm{d}\nu^+(x) = \frac{1}{\theta}\int_{\mathbb{R}} \mathrm{e}^{tx} g^+(x)\,\mathrm{d}\mu(x) = \frac{1}{\theta}\int_{\mathbb{R}} \mathrm{e}^{tx} g^-(x)\,\mathrm{d}\mu(x) = \int_{\mathbb{R}} \mathrm{e}^{tx}\,\mathrm{d}\nu^-(x).$$

Thus, $\nu^+ \equiv \nu^-$ and choosing $A^+ = \{x : g^+(x) > 0\} \subseteq \{x : g^-(x) = 0\}$, we are led to the contradiction $1 = \nu^+(A^+) = \nu^-(A^+) = 0$.

Acknowledgements

This work was partially supported by the University of Athens' Research fund under Grant No. 70/4/5637. We would like to thank Professor Roger W. Johnson for providing us with a copy of his paper. Thanks are also due to an anonymous referee for the careful reading of the manuscript and for some helpful comments. N. Papadatos is devoting this work to his six years old little daughter, Dionyssia, who is suffering from a serious health disease.

References

[1] Abbey, J.L. and David, H.T. (1970). The construction of uniformly minimum variance unbiased estimators for exponential distributions. *Ann. Math. Statist.* **41** 1217–1222. MR0285064

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

[3] Akhiezer, N.I. (1965). *The Classical Moment Problem and Some Related Questions in Analysis*. New York: Hafner Publishing Co. MR0184042

[4] Alharbi, A.A. and Shanbhag, D.N. (1996). General characterization theorems based on versions of the Chernoff inequality and the Cox representation. *J. Statist. Plann. Inference* **55** 139–150. MR1423963

[5] Asai, N., Kubo, I. and Kuo, H. (2003). Multiplicative renormalization and generating functions I. *Taiwanese J. Math.* **7** 89–101. MR1961041

[6] Asai, N., Kubo, I. and Kuo, H. (2004). Multiplicative renormalization and generating functions II. *Taiwanese J. Math.* **8** 593–628. MR2105554

[7] Beale, F.S. (1941). On a certain class of orthogonal polynomials. *Ann. Math. Statist.* **12** 97–103. MR0003852

[8] Berg, C. and Christensen, J.P.R. (1981). Density questions in the classical theory of moments. *Ann. Inst. Fourier (Grenoble)* **31** VI 99–114. MR0638619

[9] Blight, B.J.N. and Rao, P.V. (1974). The convergence of Bhattacharyya bounds. *Biometrika* **61** 137–142. MR0381097

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

[11] Cacoullos, T., Papadatos, N. and Papathanasiou, V. (2001). An application of a density transform and the local limit theorem. *Teor. Veroyatnost. i Primenen.* **46** 803–810; translation in *Theory Probab. Appl.* **46** (2003) 699–707. MR1971837

[12] Cacoullos, T. and Papathanasiou, V. (1986). Bounds for the variance of functions of random variables by orthogonal polynomials and Bhattacharyya bounds. *Statist. Probab. Lett.* **4** 21–23. MR0822720

[13] Cacoullos, T. and Papathanasiou, V. (1989). Characterizations of distributions by variance bounds. *Statist. Probab. Lett.* **7** 351–356. MR1001133

[14] Cacoullos, T., Papathanasiou, V. and Utev, S.A. (1994). Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* **22** 1607–1618. MR1303658

[15] Chen, L.H.Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3** 534–545. MR0428387

[16] Chernoff, H. (1981). A note on an inequality involving the normal distribution. *Ann. Probab.* **9** 533–535. MR0614640

[17] Diaconis, P. and Zabell, S. (1991). Closed form summation for classical distributions: Variations on a theme of De Moivre. *Statist. Sci.* **6** 284–302. MR1144242

[18] Goldstein, L. and Reinert, G. (1997). Stein's method and the zero-bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7** 935–952. MR1484792

[19] Goldstein, L. and Reinert, G. (2005). Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.* **18** 237–260. MR2132278

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

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

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

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

[24] Lefévre, C., Papathanasiou, V. and Utev, S.A. (2002). Generalized Pearson distributions and related characterization problems. *Ann. Inst. Statist. Math.* **54** 731–742. MR1954042

[25] López-Blázquez, F. and Salamanca-Miño, B. (2000). Estimation based on the winzorized mean in the geometric distribution. *Statistics* **35** 81–95. MR1820824

[26] Papadatos, N. and Papathanasiou, V. (2001). Unified variance bounds and a Stein-type identity. In *Probability and Statistical Models With Applications* (C.A. Charalambides, M.V. Koutras and N. Balakrishnan, eds.) 87–100. New York: Chapman & Hall/CRC.

[27] Papadatos, N. and Papathanasiou, V. (2002). Poisson approximation for a sum of dependent indicators: An alternative approach. *Adv. in Appl. Probab.* **34** 609–625. MR1929600

[28] Papathanasiou, V. (1995). A characterization of the Pearson system of distributions and the associated orthogonal polynomials. *Ann. Inst. Statist. Math.* **47** 171–176. MR1341214

[29] Privault, N. (2001). Extended covariance identities and inequalities. *Statist. Probab. Lett.* **55** 247–255. MR1867528

[30] Riesz, M. (1923). Sur le problème des moments et le théorème de Parseval correspondant (in French). *Acta Litt. Ac. Sci. (Szeged)* **1** 209–225.

[31] Schoutens, W. (2001). Orthogonal polynomials in Stein's method. *J. Math. Anal. Appl.* **253** 515–531. MR1808151

[32] Seth, G.R. (1949). On the variance of estimates. *Ann. Math. Statist.* **20** 1–27. MR0029141

[33] Stein, C.M. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 583–602. Berkeley, CA: Univ. California Press. MR0402873

[34] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

Received June 2008 and revised February 2010