Brazilian Journal of Probability and Statistics
2016, Vol. 30, No. 2, 171–195
DOI: 10.1214/14-BJPS271

Parametric Stein operators and variance bounds

Christophe Ley[a] **and Yvik Swan**[b]

[a]*Université libre de Bruxelles*
[b]*Université de Liège*

Abstract. Stein operators are (differential/difference) operators which arise within the so-called Stein's method for stochastic approximation. We propose a new mechanism for constructing such operators for arbitrary (continuous or discrete) parametric distributions with continuous dependence on the parameter. We provide explicit general expressions for location, scale and skewness families. We also provide a general expression for discrete distributions. We use properties of our operators to provide upper and lower variance bounds (only lower bounds in the discrete case) on functionals $h(X)$ of random variables X following parametric distributions. These bounds are expressed in terms of the first two moments of the derivatives (or differences) of h. We provide general variance bounds for location, scale and skewness families and apply our bounds to specific examples (namely the Gaussian, exponential, gamma and Poisson distributions). The results obtained via our techniques are systematically competitive with, and sometimes improve on, the best bounds available in the literature.

1 Introduction

Let g be a given target density (continuous or discrete) and let $X \sim g$. Choose a probability metric d (Kolmogorov, Wasserstein, Total Variation, ...) and suppose that we aim to estimate the distance $d(W, X)$ between the law of some random variable W and that of X. Stein's method (introduced for Gaussian approximation in Stein (1970/1971) and for Poisson approximation in Chen (1975)) is a technique initially designed for this purpose and can be broken down into three steps, namely

(A) construct a suitable differential or difference operator $f \mapsto \mathcal{T}_g(f)$ such that

$$X \sim g \quad \Longleftrightarrow \quad \mathrm{E}\big[\mathcal{T}_g(f)(X)\big] = 0 \qquad \text{for all } f \in \mathcal{F}(g),$$

with $\mathcal{F}(g)$ a specific (g-dependent) class of *test functions*;

(B) determine a subclass $\mathcal{F}_d(g) \subset \mathcal{F}(g)$ such that

$$d(W, X) = \sup_{f \in \mathcal{F}_d(g)} \big|\mathrm{E}\big[\mathcal{T}_g(f)(W)\big]\big|,$$

and determine bounds on the functions $f \in \mathcal{F}_d(g)$;

Key words and phrases. Chernoff inequality, Cramér–Rao inequality, parameter of interest, Stein characterization, Stein's method.

Received September 2013; accepted November 2014.

(C) use the knowledge about W (e.g., its distribution or that it is a sum of weakly dependent random variables, ...) in order to provide estimates on $\sup_{f\in\mathcal{F}_d(g)} |\mathrm{E}[\mathcal{T}_g(f)(W)]|$.

The bounds mentioned in (B) are sometimes called Stein factors (see, e.g., Röllin (2012), Brown and Xia (1995)) and are usually obtained by solving a "Stein equation" of the form $\mathcal{T}_g f = h$ for some h well-chosen. Although still mainly applied to Gaussian approximation (Barbour and Chen (2005), Nourdin and Peccati (2012), Chen, Goldstein and Shao (2011)) and Poisson approximation (Barbour, Holst and Janson (1992)), the method has also been proven in recent years to be very powerful for other types of approximation problems (Nourdin and Peccati (2009), Luk (1994), Picket (2004), Döbler (2012), Goldstein and Reinert (2013), Peköz, Röllin and Ross (2013), Peköz and Röllin (2011), Chatterjee, Fulman and Röllin (2011)).

The success of the method outlined above is often described as "magical", see, for example, Barbour and Chen (2014). In fact, the key lies in the exquisitely agreeable properties of the pair $(\mathcal{T}_g(\cdot), \mathcal{F}_d(g))$. There are several well-documented ways of constructing a Stein operator $\mathcal{T}_g(\cdot)$ along with the corresponding Stein class $\mathcal{F}_d(g)$; three classical constructions are (i) the *generator approach* introduced in Götze (1991), Barbour (1990), (ii) the *density approach* introduced in Stein (1986), Stein et al. (2004) and developed in Ley, Reinert and Swan (2014), and (iii) the *orthogonal polynomial approach* introduced in Diaconis and Zabell (1991) and further developed in Goldstein and Reinert (2005). Applying these techniques (or variations thereof), useful Stein operators have now been discovered for a wide variety of targets, see, for example, Götze and Tikhomirov (2003), Reinert (2004), Goldstein and Reinert (2005), Döbler (2012), Goldstein and Reinert (2013), Ley and Swan (2013a, 2013b) or the dedicated web page https://sites.google.com/site/yvikswan/about-stein-s-method for an up-to-date list of references. A handbook detailing such results is also currently in preparation, see the forthcoming Döbler et al. (2015).

Example 1.1. For instance, if $g = \phi$ is the standard Gaussian density, then a routine application of the density approach gives the first-order operator $\mathcal{T}_{0,\phi}(f)(x) = f'(x) - xf(x)$, while the generator approach yields the second-order operator $\tilde{\mathcal{T}}_{0,\phi}(f)(x) = f''(x) - xf'(x)$ and the orthogonal polynomial approach yields, among others, the collection of operators $\mathcal{T}_{n,\phi}(f)(x) = H_n(x)f'(x) - H_{n+1}(x)f(x)$, $n \geq 1$, with H_n the nth Hermite polynomial. If g is the rate-1 exponential distribution then suitable modifications of the density approach result in the operators $\mathcal{T}_{1,g}(f)(x) = -f'(x) + f(x)$ and $\mathcal{T}_{1,g}(f)(x) = -xf'(x) + (x-1)f(x)$; both have been used for exponential approximation problems (Chatterjee, Fulman and Röllin (2011), Peköz and Röllin (2011)).

Stein operators allow, in essence, to write general integration by parts formulas of the form

$$\mathrm{E}\big[f(X)h'(X)\big] = \mathrm{E}\big[\mathcal{T}_g(f)(X)h(X)\big]. \tag{1.1}$$

There are many ways to put such identities to use. For instance, setting $f = 1$ in (1.1) (if this is permitted) and applying the Cauchy–Schwarz inequality to the right-hand side we deduce that

$$\frac{(\mathrm{E}[h'(X)])^2}{\mathrm{E}[(\mathcal{T}_g(1)(X))^2]} \leq \mathrm{E}\big[(h(X))^2\big] \tag{1.2}$$

for all appropriate test functions h. This is a generalization of the celebrated Cramér–Rao inequality, with $\mathrm{E}[(\mathcal{T}_g(1)(X))^2]$ being some form of Fisher information for X. In particular if $g = \phi$ is the density of a standard Gaussian random variable then $\mathcal{T}_\phi(1)(x) = -x$ and (1.2) particularizes to $(\mathrm{E}[h'(X)])^2 \leq \mathrm{Var}[h(X)]$ (provided $\mathrm{E}[h(X)] = 0$). Chernoff (1980, 1981) used a method involving Hermite polynomials to prove that if X is Gaussian then a converse inequality also holds, yielding

$$\big(\mathrm{E}\big[h'(X)\big]\big)^2 \leq \mathrm{Var}\big[h(X)\big] \leq \mathrm{E}\big[\big(h'(X)\big)^2\big] \tag{1.3}$$

with equality on both sides if and only if h is linear. Chen presented in Chen (1982) an ingenious way of using a Gaussian version of (1.1) (namely Stein's covariance identity) to prove the bound (1.3) also in the multivariate setting. Chen's approach was rapidly seen to be robust to a change in the target distribution and Klaassen (1985) proposed a unified version of (1.3) valid under very few assumptions on X. These pioneering works spawned a stream of papers wherein similar inequalities were obtained and exploited under various assumptions on X, see, for example, Cacoullos (1982), Chernoff (1981), Chen (1982), Borovkov and Utev (1984), Cacoullos, Papathanasiou and Utev (1994), Cacoullos and Papathanasiou (1995), Houdré and Kagan (1995), Papadatos and Papathanasiou (2001), Afendras, Papadatos and Papathanasiou (2011). To put these results in a broader perspective, variance bounds are related to classical topics from functional analysis, such as concentration of measures (see, e.g., Ledoux (2001)) and Poincaré, logarithmic Sobolev and Sobolev inequalities (see Bakry, Gentil and Ledoux (2014), part II). We also refer the reader to the recent work of Ledoux, Nourdin and Peccati (2015) for a new and striking connexion between logarithmic Sobolev inequalities and Stein's method.

In this paper, we present a new way of constructing Stein operators and show how to use the resulting identities to obtain lower and upper variance bounds. We are therefore meddling with two classical topics in a seemingly classical way. Our approach is nevertheless important in at least two aspects. First, the mechanism we use is sufficiently abstract to generate a wealth of operators *and* variance bounds (some known and others new) for all matters of distributions in a uniform way. Second, our construction relies on a new parametric interpretation (in the statistical sense) of the Stein operators and of the resulting variance bounds. For instance, we show that Chernoff's bounds (1.1) ought to be read as *location-based* bounds, that is, bounds obtained by optimising with respect to μ in the location Gaussian model $\phi(\cdot - \mu)$ for $\mu \in \mathbb{R}$; we also show how to construct *scale-based* bounds by optimising with respect to σ in the scaled Gaussian model $\sigma\phi(\sigma(\cdot - \mu))$ for $\sigma \in \mathbb{R}^+$, hereby recovering the bound

$$\frac{1}{2}\big(\mathrm{E}\big[Xh'(X)\big]\big)^2 \leq \mathrm{Var}\big[h(X)\big]$$

already discussed in Cacoullos (1982), Ledoux (2001); finally we obtain *skewness-based* bounds by optimising with respect to δ in the skewed Gaussian model $(H_\delta)'(x)\phi(H_\delta(x))$, for H_δ some skewing function, obtaining in particular the bound

$$\frac{(\mathrm{E}[\sqrt{1+X^2}h'(X)])^2}{\kappa} \leq \mathrm{Var}\big[h(X)\big]$$

(for $\kappa \approx 2.34432$) which, to the best of our knowledge, is new. We can also consider alternative targets such as $X \sim t_m$, the Student distribution with m degrees of freedom, for which a routine application of our Proposition 3.1 yields the bound

$$\mathrm{Var}\big(h(X)\big) \geq \frac{m+3}{m+1}\mathrm{E}\big[h'(X)\big]^2 \tag{1.4}$$

while a routine application of our Proposition 3.2 yields

$$\mathrm{Var}\big(h(X)\big) \geq \frac{m+3}{2m}\mathrm{E}\big[Xh'(X)\big]^2 \tag{1.5}$$

in both cases for $h \in C_0^1(\mathbb{R})$. Many more similar results will be discussed in the text.

Bounds such as (1.4) and (1.5) are certainly available from other approaches such as that outlined in Klaassen (1985); however such results are in general difficult to apply to any specific choice of distribution (or at least require quite demanding computations) while ours are *immediate*. Moreover, we have good reason to believe that, when applicable, the bounds obtained by our approach are systematically good. For instance, the bounds obtained in the Gaussian case are optimal; for the Student case one can for instance compare with the corresponding bounds given in Landsman, Vanduffel and Yao (2015) (ours are better); in the exponential case we again immediately obtain good bounds by a direct application of our parametric approach, see Example 3.3; similar conclusions hold in the Poisson case, see Example 3.5.

Now it is a near trivial observation that a plethora of Stein operators is available for any given distribution: for instance replacing $f(x)$ by $xf(x)$ in the classical operator $f'(x) - xf(x)$ leads to the operator $xf'(x) + (1 - x^2)f(x)$ and, considering such standardisations in all generality, obviously leads to infinitely many more operators in a straightforward fashion. See, for example, Ley, Reinert and Swan (2014) for a thorough discussion of this approach. Most of the operators obtained in such manner are of no practical use and it still remains a mystery as to which particular operator will be of interest for applications. As a rule of thumb, it seems that only operators which bear an intuitive interpretation (as, e.g., the operators arising from the generator approach) stand a chance of being good choices for the method to work. As outlined above it seems that the operators obtained by our approach (and therefore the corresponding variance bounds) are systematically good. This is perhaps due to the fact that, even though the operators we obtain could have been derived from the density approach by a suitable pre-multiplication of $f(x)$ with some function $c(x)$ (e.g., we have used $c(x) = x$ above), they now are branded with a hitherto unsuspected parametric (and therefore statistical) interpretation. It is, at this stage, still unclear what practical implications this taxonomy might have, outside of the results presented here. We do nevertheless hope that the current paper will serve as stepping stone for research on the applications of Stein's method in (semi-parametric) statistics, perhaps along the path described in the classical papers Hudson (1978) or Liu (1994).

1.1 Outline of the paper

We develop (Section 2) a new mechanism—which we call the *parametric approach*—for building Stein operators in terms of the *parameters of interest* (location parameter, scale parameter, skewness parameter, ...) of the target distribution g. We show (Sections 2.1–2.4) that the operators $\mathcal{T}_\theta(f, g)$ indeed generalize the classical Stein operators from the literature. We then use these operators to propose (Section 3) an extension of (1.3) to a wide variety of target distributions g. Detailed specific examples are provided and discussed throughout, and lengthy proofs are deferred to the end of the paper (Section 4).

2 Parametric Stein operators

Throughout, we let $\Theta \subseteq \mathbb{R}$ be a non-empty measurable subset of $\mathbb{R}$ and say that a measurable function $g : \mathbb{R} \times \Theta \to \mathbb{R}^+$ forms a family of θ*-parametric densities* on $\mathbb{R}$ (with respect to some general σ-finite dominating measure μ) if

$$\int g(x; \theta)\, d\mu(x) = 1 \qquad \text{for all } \theta \in \Theta. \tag{2.1}$$

If in (2.1) μ is the counting measure on the integers then we further have $0 \leq g(x; \theta) \leq 1$ for all x and θ. For $\theta_0 \in \Theta$ (θ_0 has of course the same parametric nature as θ), we denote by $\mathcal{G}(\mathbb{R}, \theta_0)$ the collection of θ-parametric densities on $\mathbb{R}$ for which there exist a bounded neighborhood $\Theta_0 \subset \Theta$ of θ_0 and a μ-integrable function $h : \mathbb{R} \to \mathbb{R}^+$ such that $g(x; \theta) \leq h(x)$ over $\mathbb{R}$ for all $\theta \in \Theta_0$. Given $\theta_0 \in \Theta$ and $g \in \mathcal{G}(\mathbb{R}, \theta_0)$, we write $X \sim g(\cdot; \theta_0)$ to denote a random variable distributed according to the (absolutely continuous or discrete) probability law $x \mapsto g(x; \theta_0)$.

Definition 2.1. Let θ_0 be an interior point of Θ and let $g \in \mathcal{G}(\mathbb{R}, \theta_0)$. Define $S_\theta := \{x \in \mathbb{R} | g(x;\theta) > 0\}$ as the support of $g(\cdot;\theta)$. We define the class $\mathcal{F}(g;\theta_0)$ as the collection of functions $f : \mathbb{R} \times \Theta \to \mathbb{R}$ such that there exists Θ_0 some neighborhood of θ_0 where the following three conditions are satisfied:

(i) there exists a constant $c_f \in \mathbb{R}$ (not depending on θ) such that $\int f(x;\theta) g(x;\theta)\, d\mu(x) = c_f$ for all $\theta \in \Theta_0$;

(ii) for all $x \in S_\theta$ the mapping $\theta \mapsto f(x;\theta)g(x;\theta)$ is differentiable in the sense of distributions over Θ_0;

(iii) there exists a μ-integrable function $h : \mathbb{R} \to \mathbb{R}^+$ (possibly different for each pair f and g) such that for all $\theta \in \Theta_0$ we have $|\partial_\theta (f(x;\theta)g(x;\theta))| \le h(x)$ over $\mathbb{R}$.

We define the *Stein operator* $\mathcal{T}_{\theta_0} := \mathcal{T}_{\theta_0}(\cdot, g) : \mathcal{F}(g;\theta_0) \to \mathbb{R}^*$ as

$$\mathcal{T}_{\theta_0}(f, g)(x) = \frac{\partial_\theta (f(x;\theta)g(x;\theta))|_{\theta=\theta_0}}{g(x;\theta_0)},$$

with the convention that $1/g(x;\theta_0) = 0$ outside the support $S_{\theta_0} \subseteq \mathbb{R}$ of $g(\cdot;\theta_0)$.

Let $X \sim g(\cdot;\theta)$. The conditions imposed in Definition 2.1 bear a natural interpretation. Condition (i) imposes that all functions $f \in \mathcal{F}(g;\theta_0)$ are *pivotal functions* for the model $g(\cdot;\theta)$, in the sense that $\mathrm{E}[f(X;\theta)]$ is independent of θ. Conditions (ii) and (iii) ensure that we are permitted to interchange derivatives and integrals to get

$$0 = \frac{\partial}{\partial\theta} \mathrm{E}\big[f(X;\theta)\big] = \int_{\mathcal{X}} \frac{\partial}{\partial\theta} \big(f(x;\theta)g(x;\theta)\big)\, d\mu(x)$$

for all θ in a neighbourhood of θ_0 (see, e.g., Lehmann and Casella (1998) for more information on the conditions under which these manipulations are permitted in parametric families). Dividing and multiplying the integrand on the rhs by $g(\cdot;\theta)$ we then deduce that

$$X \sim g(\cdot;\theta) \quad \Longrightarrow \quad \mathrm{E}\big[\mathcal{T}_\theta(f, g)(X)\big] = 0 \qquad \text{for all } f \in \mathcal{F}(g;\theta_0)$$

for all $\theta \in \Theta_0$. Comparing with point (A) from the Introduction leads us to interpret $\mathcal{T}_\theta$ acting on $\mathcal{F}(g;\theta)$ as a Stein operator for $g(\cdot;\theta)$.

It remains to prove the reverse implication. This is the main result of this section. The proof is quite technical and is provided in Section 4.

Theorem 2.1 (Parametric Stein characterization). *Fix an interior point $\theta_0 \in \Theta$. Let $g \in \mathcal{G}(\mathbb{R}, \theta_0)$ and Z_θ be distributed according to $g(\cdot;\theta)$, and let X be a random variable taking values on $\mathbb{R}$. Then the following two assertions hold.*

(1) *If $X \stackrel{\mathcal{D}}{=} Z_{\theta_0}$, then $\mathrm{E}[\mathcal{T}_{\theta_0}(f, g)(X)] = 0$ for all $f \in \mathcal{F}(g;\theta_0)$.*

(2) *If the support $S_\theta := S$ of $g(\cdot;\theta)$ does not depend on θ, if $\mathrm{E}[\mathcal{T}_{\theta_0}(f, g)(X)]$ exists and if $\mathrm{E}[\mathcal{T}_{\theta_0}(f, g)(X)] = 0$ for all $f \in \mathcal{F}(g;\theta_0)$, then $X | X \in S \stackrel{\mathcal{D}}{=} Z_{\theta_0}$.*

As already mentioned in the Introduction, modern literature on probability theory is peppered with Stein operators for all manners of distributions. These have so far all been constructed through variations of either Stein's density approach, Barbour and Götze's generator approach or Diaconis and Zabell's orthogonal polynomial approach. Theorem 2.1 yields a fourth tool for constructing Stein operators; we call it the *parametric approach*. In the next sections we particularize this result to three important types of parameters, namely location, scale and skewness (in each case for absolutely continuous target distributions). As we shall see, many operators used in the literature can be labelled either as location- or scale-based. The skewness-based operators are, to the best of our knowledge, new. We will also see how to apply Theorem 2.1 in the case of general discrete distributions with continuous dependence on the parameter.

2.1 Stein operators for location models

Let the dominating measure μ be the Lebesgue measure on $\mathbb{R}$ (and write dx for $d\mu(x)$). Let $\Theta = \mathbb{R}$, fix $\nu_0 \in \mathbb{R}$ (typically one takes $\nu_0 = 0$) and consider densities of the form

$$g(x; \nu) = g_0(x - \nu), \qquad \nu \in \mathbb{R}, \tag{2.2}$$

for g_0 some positive function integrating to 1 over its support. We denote by $\mathcal{G}_{\text{loc}}$ the collection of g_0's for which ν-parametric densities of the form (2.2) belong to $\mathcal{G}(\mathbb{R}, \nu_0)$.

In the present context, condition (i) of Definition 2.1 holds naturally for test functions of the form $f(x; \nu) = f_0(x - \nu)$ for some function f_0, since in this case

$$\int_{\mathbb{R}} f(x; \nu) g(x; \nu)\, dx = \int_{\mathbb{R}} f_0(x) g_0(x)\, dx$$

is indeed independent of ν. Note that we also have

$$\partial_x \big(f_0(x - \nu) g_0(x - \nu) \big) = -\partial_\nu \big(f_0(x - \nu) g_0(x - \nu) \big) \tag{2.3}$$

for all $(x, \nu) \in \mathbb{R} \times \mathbb{R}$ (we write ∂_x and ∂_ν the weak derivatives with respect to x and ν, resp.). Conditions on f_0 under which $f(x; \nu) = f_0(x - \nu)$ satisfies conditions (i)–(iii) of Definition 2.1 are summarized in the next definition.

Definition 2.2 (Location-based Stein class). Let $g_0 \in \mathcal{G}_{\text{loc}}$. We define $\mathcal{F}_{\text{loc}}(g_0; \nu_0)$ as the collection of all $f_0 : \mathbb{R} \to \mathbb{R}$ such that (i) $\int_{\mathbb{R}} f_0(x - \nu) g_0(x - \nu)\, dx = \int_{\mathbb{R}} f_0(x) g_0(x)\, dx = c_{f_0}$ some finite constant; (ii) the mapping $x \mapsto f_0(x) g_0(x)$ is differentiable in the sense of distributions; (iii) there exists an integrable function h such that $|\partial_y (f_0(y - \nu) g_0(y - \nu))|_{y=x}| \leq h(x)$ over $\mathbb{R}$ for all $\nu \in \Theta_0$, some bounded neighborhood of ν_0.

Corollary 2.1 (Location-based Stein operator). *The conclusions of Theorem 2.1 apply to any location model of the form (2.2) with $g_0 \in \mathcal{G}_{\mathrm{loc}}$ and operator*

$$\mathcal{T}_{\nu_0;\mathrm{loc}}(f_0, g_0) : \mathbb{R} \to \mathbb{R} : x \mapsto \frac{-\partial_y (f_0(y-\nu_0) g_0(y-\nu_0))|_{y=x}}{g_0(x-\nu_0)}, \qquad (2.4)$$

for $f_0 \in \mathcal{F}_{\mathrm{loc}}(g_0; \nu_0)$ and with ∂_y the derivative in the sense of distributions with respect to y.

Example 2.1. Take $g_0(x) = \phi(x)$, the density of a $\mathcal{N}(0,1)$ random variable (which clearly belongs to $\mathcal{G}_{\mathrm{loc}}$). Then, for $\nu_0 = 0$ and any weakly differentiable function $f_0 \in \mathcal{F}_{\mathrm{loc}}(\phi; 0)$, Corollary 2.1 yields the operator

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

which shows that the usual Stein operator associated with the normal distribution is, statistically speaking, associated with the location parameter. More generally, for $n \in \mathbb{N}_0$, define recursively the sequence of polynomials $H_0(x) = 1$, $H_{n+1}(x) = -H_n'(x) + x H_n(x)$ (i.e., $H_n(x)$ is the nth Hermite polynomial) and consider functions of the form $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R} : (x, \nu) \mapsto f(x; \nu) := H_n(x-\nu) f_0(x-\nu)$, where $f_0 : \mathbb{R} \to \mathbb{R}$ is chosen such that $f \in \mathcal{F}_{\mathrm{loc}}(\phi; 0)$. Restricting the operator $\mathcal{T}_{\mathrm{loc}}(\cdot, \phi)$ to this collection of f's, we find

$$\mathcal{T}_{\mathrm{loc}}(f_0, \phi)(x) = -H_n(x) f_0'(x) + H_{n+1}(x) f_0(x), \qquad n \geq 0. \qquad (2.5)$$

This family of operators was discovered by Goldstein and Reinert (2005).

Example 2.2. Take $g_0(x) = e^{-x} \mathbb{I}_{[0,\infty)}(x)$, the rate-1 exponential density (which, as for the Gaussian, clearly belongs to $\mathcal{G}_{\mathrm{loc}}$). Again setting $\nu_0 = 0$ we get the operator

$$\mathcal{T}_{\mathrm{loc}}(f_0, \mathrm{Exp}) = (-f_0'(x) + f_0(x)) \mathbb{I}_{[0,\infty)}(x) - f_0(0) \delta_{x=0}, \qquad (2.6)$$

with $\delta_{x=0}$ the Dirac delta at $x = 0$ (recall that the derivative in (2.4) is the derivative in the sense of distributions). This was first obtained in Stein et al. (2004) and used in Chatterjee, Fulman and Röllin (2011) under the restriction $f_0(0) = 0$.

Example 2.3. If g belongs to the (continuous) *exponential family* (see Lehmann and Casella (1998) for a precise definition) then it can be easily seen that Corollary 2.1 yields the known operators discussed, for example, in Hudson (1978), Hwang (1982) or Lehmann and Casella (1998).

2.2 Stein operators for scale models

Let the dominating measure μ be the Lebesgue measure on $\mathbb{R}$ (and write dx for $d\mu(x)$). Let $\Theta = \mathbb{R}_0^+$, fix $\sigma_0 \in \Theta$ (typically one takes $\sigma_0 = 1$) and consider densities of the form

$$g(x; \sigma) = \sigma g_0(\sigma x), \qquad \sigma \in \mathbb{R}_0^+, \qquad (2.7)$$

for g_0 some positive function integrating to 1 over its support. We denote by $\mathcal{G}_{\mathrm{sca}}$ the collection of g_0's for which σ-parametric densities of the form (2.7) belong to $\mathcal{G}(\mathbb{R}, \sigma_0)$.

Condition (i) of Definition 2.1 here holds naturally for test functions of the form $f(x;\sigma) = f_0(\sigma x)$ for some function f_0 since in this case

$$\int_{\mathbb{R}} f(x;\sigma)g(x;\sigma)\,dx = \int_{\mathbb{R}} f_0(x)g_0(x)\,dx$$

is indeed independent of σ. Note that we also have the relationship

$$\partial_x\big(xf_0(\sigma x)g_0(\sigma x)\big) = \partial_\sigma\big(f_0(\sigma x)\sigma g_0(\sigma x)\big) \tag{2.8}$$

for all $(x,\sigma) \in \mathbb{R} \times \mathbb{R}_0^+$. Conditions on f_0 under which $f(x;\sigma) = f_0(\sigma x)$ satisfies conditions (i)–(iii) of Definition 2.1 are summarized in the next definition.

Definition 2.3 (Scale-based Stein class). Let $g_0 \in \mathcal{G}_{\mathrm{sca}}$. We define $\mathcal{F}_{\mathrm{sca}}(g_0;\sigma_0)$ as the collection of all $f_0 : \mathbb{R} \to \mathbb{R}$ such that (i) $\int_{\mathbb{R}} f_0(\sigma x)\sigma g_0(\sigma x)\,dx = \int_{\mathbb{R}} f_0(x)g_0(x)\,dx = c_{f_0}$ some finite constant; (ii) the mapping $x \mapsto f_0(x)g_0(x)$ is differentiable in the sense of distributions; (iii) there exists an integrable function h such that $|\partial_y(yf_0(\sigma y)g_0(\sigma y))|_{y=x}| \le h(x)$ over $\mathbb{R}$ for all $\sigma \in \Theta_0$, some bounded neighborhood of σ_0.

Corollary 2.2 (Scale-based Stein operator). *The conclusions of Theorem 2.1 apply to any scale model of the form (2.7) with $g_0 \in \mathcal{G}_{\mathrm{sca}}$ and operator*

$$\mathcal{T}_{\sigma_0;\mathrm{sca}}(f_0, g_0) : \mathbb{R} \to \mathbb{R} : x \mapsto \frac{\partial_y(yf_0(\sigma_0 y)g_0(\sigma_0 y))|_{y=x}}{\sigma_0 g_0(\sigma_0 x)},$$

for $f_0 \in \mathcal{F}_{\mathrm{sca}}(g_0;\sigma_0)$ and ∂_y the derivative in the sense of distributions with respect to y.

Example 2.4. Take $g_0(x) = \phi(x)$ the density of a $\mathcal{N}(0,1)$ (which clearly also belongs to $\mathcal{G}_{\mathrm{sca}}$), that is, this time we consider the normal with the scale parameter as parameter of interest. For $\sigma_0 = 1$ and any weakly differentiable function $f_0 \in \mathcal{F}_{\mathrm{sca}}(\phi; 1)$, Corollary 2.2 yields the operator

$$\mathcal{T}_{\mathrm{sca}}(f_0, \phi)(x) = xf_0'(x) - \big(x^2 - 1\big)f_0(x),$$

which is (up to the minus sign) a particular case of (2.5) for $n = 1$.

Example 2.5. Next take $g_0(x) = e^{-x}\mathbb{I}_{[0,\infty)}(x)$ (which also belongs to $\mathcal{G}_{\mathrm{sca}}$). Note in particular how the support $\mathbb{R}^+$ is invariant under scale change. Applying Corollary 2.2 we get the operator

$$\mathcal{T}_{\mathrm{sca}}(f_0, \mathrm{Exp})(x) = \big(xf_0'(x) - (x-1)f_0(x)\big)\mathbb{I}_{[0,\infty)}(x)$$

after setting $\sigma_0 = 1$. This scale-based operator has first been exploited in Chatterjee, Fulman and Röllin (2011). More generally, choosing g the probability density function (p.d.f.) of a gamma distribution with shape $a > 0$ we obtain

$$\mathcal{T}_{\text{sca}}(f_0, \text{Gamma})(x) = \big(xf_0'(x) - (x-a)f_0(x)\big)\mathbb{I}_{[0,\infty)}(x),$$

a variant of the gamma operator used, for example, by Nourdin and Peccati (2009).

2.3 Stein operators for skewness models

Let the dominating measure μ be the Lebesgue measure on $\mathbb{R}$ (and write dx for $d\mu(x)$). Contrarily to location and scale models which are defined in a canonical way, there exist several distinct skewness models and no canonical form of asymmetry. A popular family are the sinh–arcsinh–skew (SAS) laws of Jones and Pewsey (2009). These laws are a particular case of the construction given in Ley and Paindaveine (2010) who consider monotone increasing diffeomorphisms $H_\delta : \mathbb{R} \to \mathbb{R}$ indexed by the skewness parameter $\delta \in \mathbb{R}$ in such a way that $H_0(x) = x$ is the only odd transformation. Letting g_0 be a symmetric positive function integrating to 1 over its support, this ensures that the resulting densities

$$g(x;\delta) = (H_\delta)'(x)g_0\big(H_\delta(x)\big), \tag{2.9}$$

with $(H_\delta)'(x) = \partial_x H_\delta(x)$, are indeed skewed if δ differs from 0, value for which the initial symmetric density g_0 is retrieved. The sinh–arcsinh transformation corresponds to $H_\delta(x) = \sinh(\sinh^{-1}(x) + \delta)$. We shall call the skewed distributions (2.9) LP-densities.

For these skew distributions, let $\Theta = \mathbb{R}$, and fix $\delta_0 \in \Theta$. LP-skewness models possess densities of the form (2.9), and for a given transformation H_δ we denote by $\mathcal{G}_{\text{skew}}(H_\delta)$ the collection of g_0's for which δ-parametric densities of the form (2.9) belong to $\mathcal{G}(\mathbb{R}, \delta_0)$. In order to produce the desired operators, we however further need to add the condition that both $\delta \mapsto H_\delta(\cdot)$ and $\delta \mapsto (H_\delta)'(\cdot)$ are differentiable in the sense of distributions.

Condition (i) of Definition 2.1 here holds naturally for test functions of the form $f(x;\delta) = f_0(H_\delta(x))$; the more detailed conditions are stated in the next definition.

Definition 2.4 (LP-skewness-based Stein class). Let $g_0 \in \mathcal{G}_{\text{skew}}(H_\delta)$. We define $\mathcal{F}_{\text{skew}}(g_0; H_{\delta_0})$ as the collection of all $f_0 : \mathbb{R} \to \mathbb{R}$ such that (i) $\int_{\mathbb{R}} f_0(H_\delta(x)) \times (H_\delta)'(x)g_0(H_\delta(x))\,dx = \int_{\mathbb{R}} f_0(x)g_0(x)\,dx = c_{f_0}$ some finite constant; (ii) the mapping $x \mapsto f_0(x)g_0(x)$ is differentiable in the sense of distributions; (iii) there exists an integrable function h such that $|\partial_\delta(f_0(H_\delta(x))(H_\delta)'(x)g_0(H_\delta(x)))| \le h(x)$ over $\mathbb{R}$ for all $\delta \in \Theta_0$, some bounded neighborhood of δ_0.

Corollary 2.3 (LP-skewness-based Stein operator). *The conclusions of Theorem 2.1 apply to any LP-skewness model of the form (2.9) with $g_0 \in \mathcal{G}_{\text{skew}}(H_\delta)$ and operator*

$$\mathcal{T}_{H_{\delta_0};\text{skew}}(f_0, g_0) : \mathbb{R} \to \mathbb{R} : x \mapsto \frac{\partial_\delta(f_0(H_\delta(x))(H_\delta)'(x)g_0(H_\delta(x)))|_{\delta=\delta_0}}{(H_{\delta_0})'(x)g_0(H_{\delta_0}(x))}$$

for $f_0 \in \mathcal{F}_{\text{skew}}(g_0; H_{\delta_0})$.

Given a continuous density g_0 we define (as in Jones and Pewsey (2009)) the SAS-skew-model

$$g(x;\delta) = (1+x^2)^{-1/2} C_\delta(x) g_0(S_\delta(x)),$$

where $S_\delta(x) = \sinh(\sinh^{-1}(x) + \delta)$ and $C_\delta(x) = \cosh(\sinh^{-1}(x) + \delta)$ ($g(x;\delta)$ clearly belongs to $\mathcal{G}(\mathbb{R}, \delta_0)$ for any $\delta_0 \in \mathbb{R}$). Then we have the relationship

$$\partial_x \big(C_\delta(x) f_0(S_\delta(x)) g_0(S_\delta(x))\big) = \partial_\delta \left(f_0(S_\delta(x)) \frac{C_\delta(x)}{\sqrt{1+x^2}} g_0(S_\delta(x)) \right) \tag{2.10}$$

for all weakly differentiable functions $f_0 \in \mathcal{F}_{\text{skew}}(\phi; S_{\delta_0})$. Specifying Corollary 2.3 to this skewing mechanism, we get the operator

$$\mathcal{T}_{\text{skew}}(f_0, g_0)(x) = C_{\delta_0}(x) f_0'(S_{\delta_0}(x)) + \left(\frac{S_{\delta_0}(x)}{C_{\delta_0}(x)} + C_{\delta_0}(x) \frac{g_0'(S_{\delta_0}(x))}{g_0(S_{\delta_0}(x))} \right) f_0(S_{\delta_0}(x)).$$

Fixing $\delta_0 = 0$, the above becomes

$$\mathcal{T}_{\text{skew}}(f_0, g_0)(x) = \sqrt{1+x^2} f_0'(x) + \left(\frac{x}{\sqrt{1+x^2}} + \sqrt{1+x^2} \frac{g_0'(x)}{g_0(x)} \right) f_0(x), \tag{2.11}$$

an operator which is unlike anything we have encountered in the literature.

Example 2.6. Take $g_0 = \phi$, the standard Gaussian p.d.f. and $f_0(x) = \sqrt{1+x^2} f_1(x)$ with f_1 some suitable function in (2.11). We obtain

$$\mathcal{T}_\phi(f_1)(x) = (1+x^2) f_1'(x) - (x^3 - x) f_1(x),$$

which seems to be a new operator for the Gaussian distribution.

2.4 Discrete parametric distributions

Let the dominating measure μ be the counting measure on $\mathbb{Z}$. Let $\Theta \subset \mathbb{R}$, and fix $\theta_0 \in \Theta$. Define $\mathcal{G}_{\text{dis}}$ as the collection of θ-parametric discrete densities $g \in \mathcal{G}(\mathbb{Z}, \Theta)$ such that $g(\cdot;\theta): \mathbb{Z} \to [0,1]$ has support $S = [N] := \{0, \dots, N\}$ for some $N \in \mathbb{N}_0 \cup \{\infty\}$ not depending on θ and such that the function $\theta \mapsto g(x;\theta)$ is weakly differentiable around θ_0 at all $x \in [N]$.

Define the function $D_x^+ f$ as $D_x^+ f(x;\theta) = f(x+1;\theta) - f(x;\theta)$. It is easy to check that condition (i) of Definition 2.1 here holds for test functions of the form

$$f(x;\theta) = \frac{D_x^+(f_0(x)(g(x;\theta)/g(0;\theta)))}{g(x;\theta)}, \tag{2.12}$$

since in this case

$$\sum_{x=0}^{N} f(x;\theta) g(x;\theta) = \sum_{x=0}^{N} D_x^+ \left(f_0(x) \frac{g(x;\theta)}{g(0;\theta)} \right) = f(0)$$

for all $\theta \in \mathbb{R}$. Also note that, for f of the form (2.12), we have the relationship

$$\partial_\theta\big(f(x;\theta)g(x;\theta)\big) = D_x^+\big(f_0(x)\partial_\theta\big(g(x;\theta)/g(0;\theta)\big)\big) \tag{2.13}$$

for all $(x,\theta) \in [N] \times \mathbb{R}$.

Definition 2.5 (Discrete parametric Stein class). Let $g \in \mathcal{G}_{\mathrm{dis}}$. We define $\mathcal{F}_{\mathrm{dis}}(g;\theta_0)$ as the collection of all functions $f_0 : \mathbb{Z} \to \mathbb{R}$ such that (i) $\sum_{x=0}^N D_x^+(f_0(x)\partial_\theta(g(x;\theta)/g(0;\theta))) < \infty$ and (ii) there exists a summable function $h : \mathbb{Z} \to \mathbb{R}^+$ such that $|\Delta_x^+(f_0(x)\partial_u(g(x;u)/g(0;u))|_{u=\theta})| \le h(x)$ over $\mathbb{Z}$ for all $\theta \in \Theta_0$ some neighborhood of θ_0.

Note that here condition (ii) of Definition 2.1 is always satisfied since we use the forward difference. Moreover, for finite N, the above-mentioned sum is also finite, and we have $\sum_{x=0}^N D_x^+(f_0(x)\partial_\theta(g(x;\theta)/g(0;\theta))) = -f_0(0)$ which does not depend on θ.

Corollary 2.4 (Discrete Stein operator). *The conclusions of Theorem 2.1 apply to any discrete distribution $g \in \mathcal{G}_{\mathrm{dis}}$ with operator*

$$\mathcal{T}_{\theta_0;\mathrm{dis}}(f_0, g_0) : \mathbb{Z} \to \mathbb{R} : x \mapsto \frac{D_x^+(f_0(x)\partial_\theta(g(x;\theta)/g(0;\theta))|_{\theta=\theta_0})}{g(x;\theta_0)}$$

for $f \in \mathcal{F}_{\mathrm{dis}}(g;\theta_0)$.

Example 2.7. Take $g(x;\lambda) = e^{-\lambda}\lambda^x/x!\mathbb{I}_{\mathbb{N}}(x)$, the density of a Poisson $\mathcal{P}(\lambda)$ distribution. Clearly, g belongs to $\mathcal{G}_{\mathrm{dis}}$ for all $\lambda \in \mathbb{R}_0^+$ and its support $S = \mathbb{N}$ is independent of λ. Then, for $x \in \mathbb{N}_0$ we have $\partial_\lambda(g(x;\lambda)/g(0;\lambda))|_{\lambda=\lambda_0} = \lambda_0^{x-1}/(x-1)!$ so that

$$\mathcal{T}_{\mathrm{dis}}\big(f_0, \mathcal{P}(\lambda_0)\big)(x) = e^{\lambda_0}\bigg(f_0(x+1) - \frac{x}{\lambda_0}f_0(x)\bigg)\mathbb{I}_{\mathbb{N}}(x),$$

which is (up to the scaling factor) the usual operator for the Poisson.

Example 2.8. Take $g(x;p) = (1-p)^x p\mathbb{I}_{\mathbb{N}}(x)$, the geometric Geom$(p)$ distribution, we get

$$\mathcal{T}_{\mathrm{dis}}\big(f_0, \mathrm{Geom}(p)\big)(x) = \frac{1}{p}\bigg((x+1)f_0(x+1) - \frac{x}{1-p}f_0(x)\bigg)\mathbb{I}_{\mathbb{N}}(x).$$

Example 2.9. Finally, for the binomial Bin(n,p), we obtain the p-characterizing operator

$$\mathcal{T}_{p;\mathrm{dis}}\big(f_0, \mathrm{Bin}(n,p)\big)(x) = (1-p)^{-n-2}\bigg((n-x)f_0(x+1) - \frac{1-p}{p}xf_0(x)\bigg)\mathbb{I}_{[n]}(x).$$

These last two operators are not new, and can be obtained (up to scaling factors) as in Holmes (2004) and Ley, Reinert and Swan (2014) via the generator approach.

3 Variance bounds

Consider a θ-parametric density $g \in \mathcal{G}(\mathbb{R}, \theta_0)$ with associated Stein class $\mathcal{F}(g; \theta_0)$ and operator $\mathcal{T}_{\theta_0}(\cdot, g)$ at some point $\theta_0 \in \Theta$. Suppose, for simplicity, that the support S_θ of $g(\cdot; \theta)$ is a real interval with closure $\bar{S}_\theta = [a, b]$ for $-\infty \le a < b \le \infty$, where $a = a_\theta$ and $b = b_\theta$. (If μ is the counting measure, then $S = \{a, a + 1, \dots, b - 1, b\}$.)

We single out the subclass $\mathcal{F}_1(g; \theta_0) \subset \mathcal{F}(g; \theta_0)$ (often written simply $\mathcal{F}_1$ in the sequel) of test functions such that, for all θ in some bounded neighborhood Θ_0 of θ_0, (i) $f(x; \theta) \ge 0$ over $\mathbb{R}$, (ii) $\int_{\mathbb{R}} f(x; \theta) g(x; \theta)\, d\mu(x) = 1$ and (iii) the function

$$\tilde{f}(x; \theta) = \frac{1}{g(x; \theta)} \int_a^x \partial_\theta \big(f(y; \theta) g(y; \theta) \big)\, d\mu(y) \tag{3.1}$$

satisfies the boundary conditions

$$\tilde{f}(a; \theta) g(a; \theta) = \tilde{f}(b; \theta) g(b; \theta) = 0 \tag{3.2}$$

(interpreted as a limit if either a or b is infinite) for all $\theta \in \Theta_0$. For $f \in \mathcal{F}_1(g; \theta_0)$, the function $g^\star(x; \theta) = f(x; \theta) g(x; \theta)$ is again a θ-parametric density and we have the "exchange of derivatives" relation

$$\partial_\theta \big(f(x; \theta) g(x; \theta) \big) = \partial_x \big(\tilde{f}(x; \theta) g(x; \theta) \big) \qquad \text{for all } x \in \mathbb{R} \text{ and all } \theta \in \Theta_0. \tag{3.3}$$

See, for illustrations, equations (2.3), (2.8), (2.10) and (2.13). For ease of reference we call the pair $(f, \tilde{f})$ exchanging around θ. If μ is the counting measure, then the derivative ∂_x in (3.3) is to be replaced with the forward difference operator D_x^+.

Example 3.1. We provide details of the construction in the setting of Section 2.2. In this case the parameter θ is positive and its role is multiplicative in the sense that

$$f(x; \theta) = f_0(x\theta).$$

Then, from (2.8), we see that the pair $(f, \tilde{f})$ with

$$\tilde{f}(x; \theta) = x/\theta f_0(x\theta)$$

is exchanging around θ. It is also easily checked that (3.1) is satisfied, because $\tilde{f}(x; \theta) g_0(x\theta) = x f_0(x\theta) g(x\theta)$ and

$$\begin{aligned} \int_0^x \partial_\theta \big(\theta f_0(y\theta) g(y\theta) \big)\, dy &= \partial_\theta \int_0^x \big(\theta f_0(y\theta) g(y\theta) \big)\, dy \\ &= \partial_\theta \int_0^{x\theta} f_0(y) g(y)\, dy = x f_0(x\theta) g(x\theta). \end{aligned}$$

3.1 The continuous case

Take the dominating measure μ the Lebesgue measure (and write dx for $d\mu(x)$). All distributions considered in this section are absolutely continuous with respect to μ, and we use the superscript $'$ to indicate a (classical) strong derivative.

Our generalized variance bounds are provided in the following theorem, whose proof (given in Section 4) strongly relies on the crucial condition (3.2) and on the Stein characterizations of Theorem 2.1.

Theorem 3.1. *Let $g \in \mathcal{G}(\mathbb{R}, \theta_0)$ and $X \sim g(\cdot; \theta_0)$. Choose $f \in \mathcal{F}_1(g; \theta_0)$ and let $(f, \tilde{f})$ be exchanging around θ. Let $X^\star_{f,\theta_0} \sim g^\star(\cdot; \theta_0) = f(\cdot; \theta_0)g(\cdot; \theta_0)$. Define $\varphi_{\theta_0,g^\star}(x) := \partial_\theta(\log(g^\star(x; \theta)))|_{\theta=\theta_0}$ $(= \mathcal{T}_{\theta_0}(f, g)(x)/f(x; \theta_0))$ and $\mathcal{I}(\theta_0, g^\star) := \mathrm{E}[(\varphi_{\theta_0,g^\star}(X^\star_{f,\theta_0}))^2]$. Then*

$$\mathrm{Var}\big[h\big(X^\star_{f,\theta_0}\big)\big] \geq \frac{(\mathrm{E}[h'(X)\tilde{f}(X; \theta_0)])^2}{\mathcal{I}(\theta_0, g^\star)} \tag{3.4}$$

for all $h \in C^1_0(\mathbb{R})$. If, furthermore, $x \mapsto \varphi_{\theta_0,g^\star}(x)$ is strictly monotone and strongly differentiable over its support then

$$\mathrm{Var}\big[h\big(X^\star_{f,\theta_0}\big)\big] \leq \mathrm{E}\bigg[\frac{(h'(X))^2}{-\varphi'_{\theta_0,g^\star}(X)}\tilde{f}(X; \theta_0)\bigg] \tag{3.5}$$

for all $h \in C^1_0(\mathbb{R})$. Moreover, equality holds in (3.4) and (3.5) if and only if $h(x) \propto \varphi_{\theta_0,g^\star}(x)$ for all x.

Remark 3.1. The function $\varphi_{\theta_0,g^\star}$ is the score function of $X^\star_{f,\theta_0}$, while the quantity $\mathcal{I}(\theta_0, g^\star)$ is its Fisher information. In the sequel, we will generally not use the cumbersome indexation by $(\theta_0, g^\star)$ in the notation for the score and Fisher information of $X^\star_{f,\theta_0}$. We rather opt for more handy notation such as

$$\mathcal{I}_{\mathrm{loc}}(g), \qquad \mathcal{I}_{\mathrm{sca}}(g) \quad \text{and} \quad \mathcal{I}_{\mathrm{skew}}(g)$$

indicating the parametric nature of θ as well as the reference density g.

Remark 3.2. The upper bound in (3.5) is always positive. Indeed, first observe that if $\varphi_{\theta_0,g^\star}$ is a diffeomorphism then it is, in particular, strictly monotone over the support S_{θ_0} and the function $x \mapsto \partial_\theta(f(x; \theta)g(x; \theta))|_{\theta=\theta_0}$ changes sign exactly once (because $\int_a^b \partial_\theta(f(x; \theta)g(x; \theta))|_{\theta=\theta_0}\,dx = 0$). Hence if $\varphi_{\theta_0,g^\star}$ is monotone increasing (resp., decreasing) then $\tilde{f}(x; \theta_0) \leq 0$ (resp., $\tilde{f}(x; \theta_0) \geq 0$) for all $x \in S_{\theta_0}$ so that the upper bound in (3.5) is positive.

A natural choice of test function in Theorem 3.1 is the constant function $f(x; \theta) = 1$, for which $g^\star(x; \theta) = g(x; \theta)$ and thus $X^\star_{f,\theta_0} \stackrel{\mathcal{L}}{=} X$. This choice is not always permitted: if, for example, the support of g depends on the parameter and if the density does not cancel at the edges of the support then condition (3.2) cannot be satisfied and our proofs break down. In practice, the problem is avoided by imposing the technical assumption that the support of $g(\cdot;\theta)$ is either open or does not depend on θ. In this case the choice $f(x;\theta) = 1$ is permitted and, using (2.3), (2.8) and (2.10) (which are the specific versions of (3.3) with respect to the different roles of the parameters considered in Section 2) we obtain explicit forms for the exchanging functions $\tilde{f}$, and thus explicit forms of the variance bounds from Theorem 3.1. In the next three results, we consider a θ-parametric density $g \in \mathcal{G}(\mathbb{R}, \theta_0)$ and let $X \sim g(\cdot;\theta_0)$.

Proposition 3.1 (Location-based variance bounds). *Let $\theta = \mu \in \mathbb{R}$ be a location parameter and $g(x;\mu) = g_0(x-\mu)$ a location model for $g_0 \in C_0^1(S)$ with open support S. Then the exchanging function for $f(x;\mu) = f_0(x-\mu) \in \mathcal{F}_{\text{loc}}(g_0;\mu_0)$ around μ is $\tilde{f}(x;\mu) = -f_0(x-\mu)$. The location-score function (expressed in terms of $y = x-\mu$) is*

$$\varphi_{g_0,\text{loc}}(y) = -\frac{g_0'(y)}{g_0(y)}\mathbb{I}_S(y).$$

If $\varphi_{g_0,\text{loc}}$ is strictly monotone and strongly differentiable on S, then the location-based variance bounds read

$$\frac{(\mathrm{E}[h'(X)])^2}{\mathcal{I}_{\text{loc}}(g_0)} \le \mathrm{Var}\big[h(X)\big] \le \mathrm{E}\bigg[\frac{(h'(X))^2}{\varphi_{g_0,\text{loc}}'(X-\mu_0)}\bigg] \tag{3.6}$$

for $h \in C_0^1(\mathbb{R})$, with $\mathcal{I}_{\text{loc}}(g_0) := \mathrm{E}[(\varphi_{g_0,\text{loc}}(X-\mu_0))^2]$.

Proposition 3.2 (Scale-based variance bounds). *Let $\theta = \sigma \in \mathbb{R}_0^+$ be a scale parameter and $g(x;\sigma) = \sigma g_0(\sigma x)$ a scale model for $g_0 \in C_0^1(S)$ with either open support S or support S invariant under scale change. Then the exchanging function for $f(x;\sigma) = f_0(\sigma x) \in \mathcal{F}_{\text{sca}}(g_0;\sigma_0)$ around σ is $\tilde{f}(x;\sigma) = \frac{x}{\sigma} f_0(\sigma x)$. The scale-score function (expressed in terms of $y = \sigma x$) is*

$$\varphi_{g_0,\text{scale}}(y) = \frac{1}{\sigma}\bigg(1 + y\frac{g_0'(y)}{g_0(y)}\bigg)\mathbb{I}_S(y).$$

If $\varphi_{g_0,\text{scale}}$ is strictly monotone and strongly differentiable on S, then the scale-based variance bounds read

$$\frac{(\mathrm{E}[h'(X)X])^2}{\sigma_0^2\mathcal{I}_{\text{sca}}(g_0)} \le \mathrm{Var}\big[h(X)\big] \le \mathrm{E}\bigg[\frac{(h'(X))^2X}{-\sigma_0^2\varphi_{g_0,\text{scale}}'(\sigma_0 X)}\bigg] \tag{3.7}$$

for $h \in C_0^1(\mathbb{R})$, with $\mathcal{I}_{\text{sca}}(g_0) := \mathrm{E}[(\varphi_{g_0,\text{scale}}(\sigma_0 X))^2]$.

Proposition 3.3 (SAS-based variance bounds). *Let $\theta = \delta \in \mathbb{R}$ be a skewness parameter and $g(x;\delta) = C_\delta(x)/\sqrt{1+x^2}\,g_0(S_\delta(x))$ the SAS-skewness model for*

$g_0 \in C_0^1(S)$ with open support S. Then the exchanging function for $f(x;\sigma) = f_0(S_\delta(x)) \in \mathcal{F}_{\mathrm{skew}}(g_0; S_{\delta_0})$ around δ is $\tilde{f}(x;\delta) = \sqrt{1+x^2} f_0(S_\delta(x))$. The skewness-score function (expressed in terms of $y = S_\delta(x)$) is

$$\varphi_{g_0,\mathrm{skew}}(y) = \left(\frac{y}{C_\delta(S_\delta^{-1}(y))} + C_\delta(S_\delta^{-1}(y))\frac{g_0'(y)}{g_0(y)}\right)\mathbb{I}_S(y).$$

If $\varphi_{g_0,\mathrm{skew}}(x)$ is monotone and strongly differentiable on S, then the SAS-based variance bounds read

$$\frac{(\mathrm{E}[h'(X)\sqrt{1+X^2}])^2}{\mathcal{I}_{\mathrm{skew}}(g_0)} \le \mathrm{Var}[h(X)] \le \mathrm{E}\left[\frac{(h'(X))^2\sqrt{1+X^2}}{-C_{\delta_0}(X)\varphi'_{g_0,\mathrm{skew}}(S_{\delta_0}(X))}\right] \tag{3.8}$$

for $h \in C_0^1(\mathbb{R})$, with $\mathcal{I}_{\mathrm{skew}}(g_0) := \mathrm{E}[(\varphi_{g_0,\mathrm{skew}}(S_{\delta_0}(X)))^2]$.

The lower bounds in (3.6), (3.7) and (3.8) hold without condition on the monotonicity of the score function. In all cases the bounds are tight, in the sense that equality holds if and only if the test function h is proportional to the score function.

In what follows, we shall apply Propositions 3.1 to 3.3 to three examples of probability laws, namely the Gaussian, the exponential and the gamma. We consider all three examples as location–scale models, but we apply the SAS-skewing mechanism only to the Gaussian distribution (as the others are already skewed over $\mathbb{R}$).

Example 3.2. Once again take $g_0(x) = \phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ the standard Gaussian density. Then, of course, $g_0'(x)/g_0(x) = -x$ and $f = 1$ belongs to $\mathcal{F}_1$ for any type of parameter. Applying the propositions for $\mu_0 = 0$ (location case), $\sigma_0 = \sigma$ (scale case) and $\delta_0 = 0$ (skewness case) we get

$$\varphi_{\phi,\mathrm{loc}}(x) = x, \qquad \varphi_{\phi,\mathrm{sca}}(x) = \frac{1}{\sigma}(1-x^2) \quad \text{and} \quad \varphi_{\phi,\mathrm{skew}}(x) = \frac{-x^3}{\sqrt{1+x^2}}.$$

Only the location score function is a "sensible" diffeomorphism (indeed, the derivative of the skewness score vanishes at the origin, leading to an infinite upper bound). Simple computations yield

$$\mathcal{I}_{\mathrm{loc}}(\phi) = 1, \qquad \mathcal{I}_{\mathrm{sca}}(\phi) = \frac{2}{\sigma^2} \quad \text{and}$$
$$\mathcal{I}_{\mathrm{skew}}(\phi) = 3 - \sqrt{\frac{e\pi}{2}}\,\mathrm{Erfc}(1/\sqrt{2}) =: \kappa \approx 2.34432.$$

We thus sequentially obtain the location-based variance bounds

$$(\mathrm{E}[h'(X)])^2 \le \mathrm{Var}[h(X)] \le \mathrm{E}[(h'(X))^2],$$

with equality if and only if h is linear (this is the well-known bound (1.3); moreover, adding a scale parameter σ in this location setting results in dividing both the upper and lower bound by σ^2) as well as the scale-based bound

$$\frac{1}{2}\big(\mathrm{E}[Xh'(X)]\big)^2 \le \mathrm{Var}[h(X)]$$

with equality if and only if $h(x) \propto 1-x^2$ (this bound is given in Cacoullos (1982), Klaassen (1985) and Ledoux (2001)) and also the skewness-based bound

$$\frac{(\mathrm{E}[\sqrt{1+X^2}h'(X)])^2}{\kappa} \le \mathrm{Var}[h(X)]$$

with equality if and only if $h(x) \propto x^3/\sqrt{1+x^2}$. This last bound seems new.

Example 3.3. Take $g_0(x) = e^{-x}\mathbb{I}_{[0,\infty)}(x)$, the rate-1 exponential density; here $f = 1$ is only permitted in the scale case and we have $g_0'(x)/g_0(x) = -1$ (for $x > 0$). Applying the propositions for $\sigma_0 = \lambda$ we get

$$\varphi_{\mathrm{Exp,sca}}(x) = \frac{1}{\lambda}(1-x)\mathbb{I}_{[0,\infty)}(x).$$

This scale-score function is clearly a diffeomorphism. Also $\mathcal{I}_{\mathrm{sca}}(\mathrm{Exp}) = \frac{1}{\lambda^2}$, which yields the scale-based variance bounds

$$\big(\mathrm{E}[Xh'(X)]\big)^2 \le \mathrm{Var}[h(X)] \le \frac{1}{\lambda}\mathrm{E}\big[X\big(h'(X)\big)^2\big]; \tag{3.9}$$

the upper bound was previously obtained in Ledoux (2001), (5.18). For the sake of comparison, Cacoullos (1982), Proposition 4.3, proposes the lower and upper bounds

$$\big(\mathrm{E}[Xh'(X)]\big)^2 \le \mathrm{Var}[h(X)] \le \frac{1}{\lambda^2}\mathrm{Var}[h'(X)] + \frac{1}{\lambda}\mathrm{E}\big[X\big(h'(X)\big)^2\big]; \tag{3.10}$$

while Klaassen (1985) proposes

$$\big(\mathrm{E}[Xh'(X)]\big)^2 \le \mathrm{Var}[h(X)] \le \frac{4}{\lambda^2}\mathrm{E}\big[\big(h'(X)\big)^2\big]. \tag{3.11}$$

The lower bound in both these seminal papers concurs with ours from (3.9). Our upper bound is evidently a strict improvement on (3.10). It also improves on (3.11) in several cases. Indeed, a simple integration by parts in our upper bound (provided that $h \in C_0^2(\mathbb{R})$) allows to rewrite it under the form

$$\frac{1}{\lambda^2}\big(\mathrm{E}\big[\big(h'(X)\big)^2\big] + 2\mathrm{E}[Xh'(X)h''(X)]\big).$$

Whenever the second term is zero (e.g., for $h(x) = x$) or negative (e.g., for $h(x) = \sqrt{x}$), our bound is better than (3.11).

Example 3.4. Finally take $g_0(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x} \mathbb{I}_{[0,\infty)}(x)$ the p.d.f. of a gamma distribution with shape $a > 0$. Here $f = 1$ is permitted in both location and scale cases if $a > 1$ and reserved to the scale case for $a \le 1$. For the sake of clarity we will only consider the case $a > 1$. We have $g_0'(x)/g_0(x) = \frac{(a-1-x)}{x}$. Applying the propositions under the respective restrictions on a and for $\mu_0 = 0$ (location case) and $\sigma_0 = b$ (scale case), we get

$$\varphi_{\mathrm{Ga,loc}}(x) = \frac{-a+1+x}{x} \mathbb{I}_{[0,\infty)}(x) \quad \text{and} \quad \varphi_{\mathrm{Ga,sca}}(x) = \frac{1}{b}(a-x) \mathbb{I}_{[0,\infty)}(x).$$

Both score functions are diffeomorphisms (on $\mathbb{R}_0^+$). Also

$$\mathcal{I}_{\mathrm{loc}}(\mathrm{Gamma}) = \begin{cases} \dfrac{1}{a-2}, & \text{if } a > 2, \\ \infty, & \text{if } 1 < a \le 2, \end{cases} \quad \text{and} \quad \mathcal{I}_{\mathrm{sca}}(\mathrm{Gamma}) = \frac{a}{b^2}.$$

This yields the following: location-based bounds

$$(a-2)\big(\mathrm{E}[h'(X)]\big)^2 \le \mathrm{Var}[h(X)] \le \frac{1}{a-1} \mathrm{E}\big[(h'(X))^2 X^2\big] \tag{3.12}$$

and scale-based bounds

$$\frac{1}{a}\big(\mathrm{E}[X h'(X)]\big)^2 \le \mathrm{Var}[h(X)] \le \frac{1}{b} \mathrm{E}\big[X (h'(X))^2\big]. \tag{3.13}$$

On the one hand Cacoullos (1982) only proposes a lower bound (which concurs with ours). On the other hand, Klaassen (1985) proposes for $a > 2$

$$\begin{aligned} \max\left(\frac{a-2}{b^2}\big(\mathrm{E}[h'(X)]\big)^2, \frac{1}{a}\big(\mathrm{E}[X h'(X)]\big)^2\right) &\le \mathrm{Var}[h(X)] \\ &\le \frac{1}{b} \mathrm{E}\big[X (h'(X))^2\big]. \end{aligned} \tag{3.14}$$

The upper bound coincides with that in (3.13), while both candidates for the lower bounds are given in (3.12) and (3.13), respectively (for a true comparison, we need to add a scale parameter in the lower location bound (3.12), resulting in a division by b^2).

We conclude this section by determining conditions on g and θ for which the bound (3.5) takes on the form

$$\mathrm{Var}\big(h(X)\big) \le d \mathrm{E}\big[(h'(X))^2\big] \tag{3.15}$$

for some positive constant d (a similar question is already addressed, in similar conditions, in Klaassen (1985)). If the special case $f = 1$ is admissible then, trivially, the constant $d = d_{g,\theta_0} = \sup_{x \in S}(-\tilde{f}(x;\theta_0)/\varphi'_{\theta_0,g^\star}(x))$ plays the required role, and the question becomes that of determining conditions under which this constant is finite. Specializing to the case of a location model we obtain the following intuitive sufficient condition.

Proposition 3.4. *Let g be a continuous density with open support and let $X \sim g$. If the function $x \mapsto (\log g(x))'$ is strict monotone decreasing and if there exists $\varepsilon > 0$ such that $-(\log g(x))'' \geq \varepsilon > 0$, then (3.15) holds with $d_{g,\mu_0} = \frac{1}{\varepsilon}$.*

Proof. Take a location model $g(x;\mu) = g(x-\mu)$ with constant test function $f(x;\mu) = 1$. Then $\tilde{f}(x;\mu) = -1$ and we compute

$$\begin{aligned}\frac{\tilde{f}(x;\mu_0)}{-\varphi'_{\mu_0,g^\star}(x)} &= \frac{1}{-g''(x-\mu_0)/(g(x-\mu_0)) + (g'(x-\mu_0)/(g(x-\mu_0)))^2} \\ &= \frac{1}{-(\log g(x-\mu_0))''}.\end{aligned}$$

The conclusion follows. □

Note that the assumptions of Proposition 3.4 hold if $g(x) = e^{-\psi(x)}$ for $\psi(x)$ a strict convex function, that is, if g is strongly unimodal on $\mathbb{R}$. We hereby recover Lemma 2.1 from Klaassen (1985). In particular, if $g(x) = (2\pi\sigma^2)^{-1/2}e^{-x^2/(2\sigma^2)}$ is the $\mathcal{N}(0,\sigma^2)$ then $\varepsilon = 1/\sigma^2$ and we reobtain the well-known upper bound $\mathrm{Var}(h(X)) \leq \sigma^2 \mathrm{E}[(h'(X))^2]$.

3.2 The discrete case

Take as dominating measure μ the counting measure. For f and g two functions such that $\sum_{x=a}^{b} D_x^+(f(x)g(x)) < \infty$ and $f(b+1)g(b+1) = f(a)g(a) = 0$, we have the discrete integration by parts formula

$$\sum_{x=a}^{b} \big(D_x^+(f(x))\big)g(x+1) = -\sum_{x=a}^{b} f(x)\big(D_x^+(g(x))\big).$$

The boundary condition (3.2) therefore allows us to deduce the following partial discrete counterpart to Theorem 3.1, whose proof is left to the reader.

Theorem 3.2. *Let $g \in \mathcal{G}(\mathbb{Z},\theta_0)$ and $X \sim g(\cdot;\theta_0)$. Choose $f \in \mathcal{F}_1(g;\theta_0)$ and let $(f,\tilde{f})$ be exchanging around θ. Let $X^\star_{f,\theta_0} \sim g^\star(\cdot;\theta_0) = f(\cdot;\theta_0)g(\cdot;\theta_0)$ and define $\varphi_{\theta_0,g^\star}(x) := \partial_\theta(\log(g^\star(x;\theta)))|_{\theta=\theta_0}$ $(= \mathcal{T}_{\theta_0}(f,g)(x)/f(x;\theta_0))$ the score function of $X^\star_{f,\theta_0}$ and $\mathcal{I}(\theta_0,g^\star) := \mathrm{E}[(\varphi_{\theta_0,g^\star}(X^\star_{f,\theta_0}))^2]$ its Fisher information. Then*

$$\mathrm{Var}\big[h\big(X^\star_{f,\theta_0}\big)\big] \geq \frac{(\mathrm{E}[D_x^-(h(X))\tilde{f}(X;\theta_0)])^2}{\mathcal{I}(\theta_0,g^\star)} \tag{3.16}$$

for all h with equality if and only if $h(x) \propto \varphi_{\theta_0,g^\star}(x)$.

Example 3.5. Take $g(x;\lambda) = e^{-\lambda}\lambda^x/x!\mathbb{I}_{\mathbb{N}}(x)$ the p.d.f. of the Poisson distribution. Then we have $\partial_\lambda g(x;\lambda) = -D_x^+(\frac{x}{\lambda}g(x;\lambda))$; in particular $1 \in \mathcal{F}_1$ because

$\tilde{1}(x;\lambda)g(x;\lambda) = \frac{x}{\lambda}g(x;\lambda)$ indeed cancels at the edges of the support of g. Also we compute $\varphi_{\lambda,g}(x) = (-1 + \frac{x}{\lambda})\mathbb{I}_{\mathbb{N}}(x)$ and $\mathcal{I}(\lambda, g) = 1/\lambda$. Applying (3.16) we conclude

$$\mathrm{Var}\big[h(X)\big] \geq \frac{1}{\lambda}\big(\mathrm{E}\big[XD_x^-\big(h(X)\big)\big]\big)^2, \tag{3.17}$$

with equality if and only if $h(x) \propto -1 + x/\lambda$ on $\mathbb{N}$. Further, using Chen's identity for the Poisson we have

$$\mathrm{E}\big[XD_x^-\big(h(X)\big)\big] = \lambda\mathrm{E}\big[D_x^+\big(h(X)\big)\big]$$

so that (3.17) is equivalent to

$$\mathrm{Var}\big[h(X)\big] \geq \lambda\big(\mathrm{E}\big[D_x^+\big(h(X)\big)\big]\big)^2 \tag{3.18}$$

given in Cacoullos (1982), Theorem 5.1, and also appearing in Klaassen (1985).

4 Proofs

Proof of Theorem 2.1. (1) Since condition (iii) allows for differentiating w.r.t. θ under the integral in condition (i) and since differentiating w.r.t. θ is allowed thanks to condition (ii), the claim follows immediately.

(2) We prove the claim in the continuous case (and write dx for $d\mu(x)$). The discrete case follows exactly along the same lines. Define, for $A \subseteq \mathbb{R}$, the mapping

$$f_A : \mathbb{R} \times \Theta_0 \to \mathbb{R} : (x, \theta) \mapsto \frac{1}{g(x;\theta)} \int_{\theta_0}^{\theta} l_A(x;u) g(x;u)\, du$$

with $l_A(x;u) := (\mathbb{I}_A(x) - \mathrm{P}(Z_u \in A))\mathbb{I}_S(x)$, where $\mathrm{P}(Z_u \in B) = \int_{\mathbb{R}} \mathbb{I}_B(x)g(x; u)\, dx$ for $B \subseteq \mathbb{R}$. Note that $\mathrm{P}(Z_u \in S) = 1$ for all $u \in \Theta_0$, since the support does not depend on the parameter of interest. We claim that f_A belongs to $\mathcal{F}(g;\theta_0)$. If this holds true the conclusion follows since then, by hypothesis,

$$\mathrm{E}\big[\mathcal{T}_{\theta_0}(f_A, g)(X)\big] = \mathrm{E}\big[l_A(X;\theta_0)\big] = \mathrm{E}\big[\mathbb{I}_{A\cap S}(X) - \mathrm{P}(Z_{\theta_0} \in A)\mathbb{I}_S(X)\big] = 0$$

and thus

$$\mathrm{P}(X \in A | X \in S) = \mathrm{P}(Z_{\theta_0} \in A)$$

for all measurable $A \subset \mathbb{R}$.

To prove the claim, first note that

$$\int_{\mathbb{R}} f_A(x;\theta)g(x;\theta)\, dx = \int_{\theta_0}^{\theta} \int_S l_A(x;u)g(x;u)\, dx\, du$$

by Fubini's theorem, which can be applied for all $\theta \in \Theta_0$ since in this case there exists a constant M such that

$$\int_{\mathbb{R}} \mathbb{I}_{(\theta_0,\theta)}(u) \int_S \big|l_A(x;u)\big| g(x;u)\, dx\, du \leq |\theta - \theta_0| \leq M$$

for all $\theta \in \Theta_0$. We also have, by definition of l_A, that

$$\int_S l_A(x; u) g(x; u)\, dx = \mathrm{P}(Z_u \in A \cap S) - \mathrm{P}(Z_u \in A)\mathrm{P}(Z_u \in S) \\ = 0.$$

Hence, f_A satisfies condition (i). Condition (ii) is easily checked. Regarding condition (iii), one sees that $\partial_t(f_A(x; t)g(x; t))|_{t=\theta} = l_A(x; \theta)g(x; \theta)$. By boundedness of the function $l_A(\cdot; \theta)$ and by definition of the class $\mathcal{G}(\mathbb{R}, \theta_0)$ we know that $|l_A(x; \theta)g(x; \theta)|$ can be bounded by an integrable function $h(x)$ uniformly in $\theta \in \Theta_0$. Hence, f_A satisfies condition (iii). We have thus proved that $f_A \in \mathcal{F}(g; \theta_0)$, and the conclusion follows. □

Proof of Theorem 3.1. For the sake of readability, throughout the proof we simply write $X^\star := X^\star_{f,\theta_0}$ and $\varphi(x) := \varphi_{\theta_0, g^\star}(x)$.

We first prove the lower bound (3.4). Take $f \in \mathcal{F}_1(g; \theta_0)$. Using (3.3) and the different assumptions (which are tailored for the following to hold) we get, on the one hand

$$\begin{aligned} \mathrm{E}\big[h(X)\mathcal{T}_{\theta_0}(f, g)(X)\big] &= \int_a^b h(x)\partial_\theta\big(f(x; \theta)g(x; \theta)\big)\Big|_{\theta_0} dx \\ &= \int_a^b h(x)\partial_x\big(\tilde{f}(x; \theta_0)g(x; \theta_0)\big)\, dx \\ &= -\int_a^b h'(x)\tilde{f}(x; \theta_0)g(x; \theta_0)\, dx = -\mathrm{E}\big[h'(X)\tilde{f}(X; \theta_0)\big] \end{aligned}$$

and, on the other hand (recall that $\mathcal{T}_{\theta_0}(f, g)(x) = \varphi(x)f(x; \theta_0)$),

$$\begin{aligned} &\big|\mathrm{E}\big[h(X)\mathcal{T}_{\theta_0}(f, g)(X)\big]\big| \\ &\quad = \big|\mathrm{E}\big[\big(h(X) - \mathrm{E}\big[h(X^\star)\big]\big)\mathcal{T}_{\theta_0}(f, g)(X)\big]\big| && (4.1) \\ &\quad \le \mathrm{E}\big[\big|h(X) - \mathrm{E}\big[h(X^\star)\big]\big|\big|\varphi(X)\big|f(X; \theta_0)\big] \\ &\quad \le \sqrt{\mathrm{E}\big[\big(h(X) - \mathrm{E}\big[h(X^\star)\big]\big)^2 f(X; \theta_0)\big]\mathrm{E}\big[f(X; \theta_0)\big(\varphi(X)\big)^2\big]} && (4.2) \\ &\quad = \sqrt{\mathrm{Var}\big[h(X^\star)\big]\mathcal{I}(\theta_0, g^\star)}, \end{aligned}$$

where (4.1) follows from the Stein characterization of Theorem 2.1 and (4.2) from the Cauchy–Schwarz inequality (recall that f is positive).

We now prove the upper bound (3.5) in the case where φ is strict monotone decreasing, the increasing case being proved exactly in the same way. Let $\varphi^{-1}(x)$ denote the inverse function of φ. Then direct manipulations involving the Cauchy–

Schwarz inequality yield

$$\begin{aligned}
\mathrm{Var}\big[h(X^\star)\big] &= \mathrm{Var}\Big[\int_0^{\varphi(X^\star)} (h\circ\varphi^{-1})'(u)\,du\Big] \le \mathrm{E}\Big[\Big(\int_0^{\varphi(X^\star)} (h\circ\varphi^{-1})'(u)\,du\Big)^2\Big] \\
&\le \mathrm{E}\Big[\int_0^{\varphi(X^\star)} 1^2\,du \int_0^{\varphi(X^\star)} \big((h\circ\varphi^{-1})'(u)\big)^2\,du\Big] \\
&= \mathrm{E}\Big[\varphi(X^\star)\int_0^{\varphi(X^\star)} \Big(\frac{h'(\varphi^{-1}(u))}{\varphi'(\varphi^{-1}(u))}\Big)^2\,du\Big].
\end{aligned}$$

Note how the latter expression is always positive: negative values of $\varphi(X^\star)$ are multiplied by a negative integral (since a positive function is integrated over $(0,\varphi(X^\star))$). Now let x_0 be the unique point in (a,b) such that $\varphi(x_0)=0$ and let $\varphi(a)=P^+$ and $\varphi(b)=-P^-$ for some $P^\pm\in\mathbb{R}\cup\{\pm\infty\}$. Then, pursuing the above,

$$\begin{aligned}
\mathrm{Var}\big[h(X^\star)\big] \le &\int_a^{x_0}\int_0^{\varphi(x)} \partial_\theta\big(f(x;\theta)g(x;\theta)\big)\Big|_{\theta=\theta_0}\Big(\frac{h'(\varphi^{-1}(u))}{\varphi'(\varphi^{-1}(u))}\Big)^2\,du\,dx \\
&+\int_{x_0}^{b}\int_0^{\varphi(x)} \partial_\theta\big(f(x;\theta)g(x;\theta)\big)\Big|_{\theta=\theta_0}\Big(\frac{h'(\varphi^{-1}(u))}{\varphi'(\varphi^{-1}(u))}\Big)^2\,du\,dx.
\end{aligned}$$

Using Fubini (which is possible since all quantities involved are positive), we deduce

$$\begin{aligned}
\mathrm{Var}\big[h(X^\star)\big] \le &\int_0^{P^+}\Big(\frac{h'(\varphi^{-1}(u))}{\varphi'(\varphi^{-1}(u))}\Big)^2\Big(\int_a^{\varphi^{-1}(u)} \partial_\theta\big(f(x;\theta)g(x;\theta)\big)\Big|_{\theta=\theta_0}dx\Big)\,du \\
&-\int_{-P^-}^{0}\Big(\frac{h'(\varphi^{-1}(u))}{\varphi'(\varphi^{-1}(u))}\Big)^2\Big(\int_{\varphi^{-1}(u)}^{b} \partial_\theta\big(f(x;\theta)g(x;\theta)\big)\Big|_{\theta=\theta_0}dx\Big)\,du.
\end{aligned}$$

From (3.3), we then get

$$\begin{aligned}
\mathrm{Var}\big[h(X^\star)\big] \le &\int_0^{P^+}\Big(\frac{h'(\varphi^{-1}(u))}{\varphi'(\varphi^{-1}(u))}\Big)^2\Big(\int_a^{\varphi^{-1}(u)} \partial_x\big(\tilde f(x;\theta_0)g(x;\theta_0)\big)\,dx\Big)\,du \\
&-\int_{-P^-}^{0}\Big(\frac{h'(\varphi^{-1}(u))}{\varphi'(\varphi^{-1}(u))}\Big)^2\Big(\int_{\varphi^{-1}(u)}^{b} \partial_x\big(\tilde f(x;\theta_0)g(x;\theta_0)\big)\,dx\Big)\,du \\
= &\int_0^{P^+}\Big(\frac{h'(\varphi^{-1}(u))}{\varphi'(\varphi^{-1}(u))}\Big)^2\tilde f\big(\varphi^{-1}(u);\theta_0\big)g\big(\varphi^{-1}(u);\theta_0\big)\,du \\
&+\int_{-P^-}^{0}\Big(\frac{h'(\varphi^{-1}(u))}{\varphi'(\varphi^{-1}(u))}\Big)^2\tilde f\big(\varphi^{-1}(u);\theta_0\big)g\big(\varphi^{-1}(u);\theta_0\big)\,du.
\end{aligned}$$

Setting $y=\varphi^{-1}(u)$ in the above and changing variables accordingly we obtain

$$\mathrm{Var}\big[h(X^\star)\big] \le \int_b^a \frac{(h'(y))^2}{\varphi'(y)}\tilde f(y;\theta_0)g(y;\theta_0)\,dy = \mathrm{E}\Big[\frac{(h'(X))^2}{-\varphi'(X)}\tilde f(X;\theta_0)\Big],$$

which is the claim. □

Acknowledgments

Christophe Ley, who is also a member of ECARES, thanks the Fonds National de la Recherche Scientifique, Communauté Française de Belgique, for support via a Mandat de Chargé de Recherche. Yvik Swan gratefully acknowledges support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy). The authors wish to sincerely thank the referees for their careful readings which led to substantial improvements of the paper.

References

Afendras, G., Papadatos, N. and Papathanasiou, V. (2011). An extended Stein-type covariance identity for the Pearson family with applications to lower variance bounds. *Bernoulli* **17**, 507–529. MR2787602

Bakry, D., Gentil, I. and Ledoux, M. (2014). *Analysis and Geometry of Markov Diffusion Operators*. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **348**. Cham: Springer. MR3155209

Barbour, A. D. (1990). Stein's method for diffusion approximations. *Probability Theory and Related Fields* **84**, 297–322. MR1035659

Barbour, A. D. and Chen, L. H. (2014). Steins (magic) method. Preprint. Available at arXiv:1411.1179.

Barbour, A. D. and Chen, L. H. Y. (2005). *An Introduction to Stein's Method*. *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.* **4**. Singapore: Singapore Univ. Press.

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

Borovkov, A. and Utev, S. (1984). On an inequality and a related characterization of the normal distribution. *Theory of Probability and Its Applications* **28**, 219–228. MR0700206

Brown, T. C. and Xia, A. (1995). On Stein–Chen factors for Poisson approximation. *Statistics and Probability Letters* **23**, 327–332. MR1342742

Cacoullos, T. (1982). On upper and lower bounds for the variance of a function of a random variable. *The Annals of Probability* **10**, 799–809. MR0659549

Cacoullos, T. and Papathanasiou, V. (1995). A generalization of covariance identity and related characterizations. *Mathematical Methods of Statistics* **4**, 106–113. MR1324694

Cacoullos, T., Papathanasiou, V. and Utev, S. A. (1994). Variational inequalities with examples and an application to the central limit theorem. *The Annals of Probability* **22**, 1607–1618. MR1303658

Chatterjee, S., Fulman, J. and Röllin, A. (2011). Exponential approximation by exchangeable pairs and spectral graph theory. *ALEA Latin American Journal of Probability and Mathematical Statistics* **8**, 1–27. MR2802856

Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *The Annals of Probability* **3**, 534–545. MR0428387

Chen, L. H. Y. (1982). An inequality for the multivariate normal distribution. *Journal of Multivariate Analysis* **12**, 306–315. MR0661566

Chen, L. H. Y., Goldstein, L. and Shao, Q.-M. (2011). *Normal Approximation by Stein's Method*. *Probability and Its Applications (New York)*. Heidelberg: Springer. MR2732624

Chernoff, H. (1980). The identification of an element of a large population in the presence of noise. *The Annals of Statistics* **8**, 1179–1197. MR0594637

Chernoff, H. (1981). A note on an inequality involving the normal distribution. *The Annals of Probability* **9**, 533–535. MR0614640

Diaconis, P. and Zabell, S. (1991). Closed form summation for classical distributions: Variations on a theme of de Moivre. *Statistical Science* **6**, 284–302. MR1144242

Döbler, C. (2012). Stein's method of exchangeable pairs for absolutely continuous, univariate distributions with applications to the Polya urn model. Preprint. Available at arXiv:1207.0533.

Döbler, C., Gaunt, R. E., Ley, C., Reinert, G. and Swan, Y. (2015). *A Handbook of Stein Operators.* In progress.

Goldstein, L. and Reinert, G. (2005). Distributional transformations, orthogonal polynomials, and Stein characterizations. *Journal of Theoretical Probability* **18**, 237–260. MR2132278

Goldstein, L. and Reinert, G. (2013). Stein's method for the Beta distribution and the Pólya–Eggenberger urn. *Journal of Applied Probability* **50**, 1187–1205. MR3161381

Götze, F. (1991). On the rate of convergence in the multivariate CLT. *The Annals of Probability* **19**, 724–739. MR1106283

Götze, F. and Tikhomirov, A. N. (2003). Rate of convergence to the semi-circular law. *Probability Theory and Related Fields* **127**, 228–276. MR2013983

Holmes, S. (2004). Stein's method for birth and death chains. In *Stein's Method: Expository Lectures and Applications*. *IMS Lecture Notes Monogr. Ser.* **46**, 42–65. Beachwood, OH: Institute of Mathematical Statistics. MR2118602

Houdré, C. and Kagan, A. (1995). Variance inequalities for functions of Gaussian variables. *Journal of Theoretical Probability* **8**, 23–30. MR1308667

Hudson, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *The Annals of Statistics* **6**, 473–484. MR0467991

Hwang, J. T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. *The Annals of Statistics* **10**, 857–867. MR0663437

Jones, M. C. and Pewsey, A. (2009). Sinh–arcsinh distributions. *Biometrika* **96**, 761–780. MR2564489

Klaassen, C. A. J. (1985). On an inequality of Chernoff. *The Annals of Probability* **13**, 966–974. MR0799431

Landsman, Z., Vanduffel, S. and Yao, J. (2015). Some Stein-type inequalities for multivariate elliptical distributions and applications. *Statistics and Probability Letters* **97**, 54–62. MR3299751

Ledoux, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Providence, RI: American Mathematical Society. MR1849347

Ledoux, M., Nourdin, I. and Peccati, G. (2015). Stein's method, logarithmic Sobolev and transport inequalities. *Geometric and Functional Analysis* **25**, 256–306. MR3320893

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. New York: Springer. MR1639875

Ley, C. and Paindaveine, D. (2010). Multivariate skewing mechanisms: A unified perspective based on the transformation approach. *Statistics and Probability Letters* **80**, 1685–1694. MR2734229

Ley, C., Reinert, G. and Swan, Y. (2014). Approximate computation of expectations: A canonical Stein operator. Preprint. Available at arXiv:1408.2998v2.

Ley, C. and Swan, Y. (2013a). Local Pinsker inequalities via Stein's discrete density approach. *IEEE Transactions on Information Theory* **59**, 5584–4491. MR3096942

Ley, C. and Swan, Y. (2013b). Stein's density approach and information inequalities. *Electronic Communications in Probability* **18**, 1–14. MR3019670

Liu, J. S. (1994). Siegel's formula via Stein's identities. *Statistics and Probability Letters* **21**, 247–251. MR1310101

Luk, H. M. (1994). Stein's method for the gamma distribution and related statistical applications. Ph.D. thesis, Univ. Southern California. MR2693204

Nourdin, I. and Peccati, G. (2009). Stein's method on Wiener chaos. *Probability Theory and Related Fields* **145**, 75–118. MR2520122

Nourdin, I. and Peccati, G. (2012). *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*. *Cambridge Tracts in Mathematics*. Cambridge: Cambridge Univ. Press. MR2962301

Papadatos, N. and Papathanasiou, V. (2001). Unified variance bounds and a Stein-type identity. In *Probability and Statistical Models with Applications* 87–100. Boca Raton, FL: Chapman & Hall/CRC.

Peköz, E. and Röllin, A. (2011). New rates for exponential approximation and the theorems of Rényi and Yaglom. *The Annals of Probability* **39**, 587–608. MR2789507

Peköz, E., Röllin, A. and Ross, N. (2013). Degree asymptotics with rates for preferential attachment random graphs. *The Annals of Applied Probability* **23**, 1188–1218. MR3076682

Picket, A. (2004). Rates of convergence of χ^2 approximations via Stein's method. Ph.D. thesis, Lincoln College, Univ. Oxford.

Reinert, G. (2004). Three general approaches to Stein's method. In *An Introduction to Stein's Method*. *Lecture Notes Series, Institute for Mathematical Sciences, National Univ. Singapore* **4**. Singapore: World Scientific.

Röllin, A. (2012). On the optimality of Stein factors. In *Probability Approximations and Beyond* 61–72. New York: Springer.

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

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

Stein, C., Diaconis, P., Holmes, S. and Reinert, G. (2004). Use of exchangeable pairs in the analysis of simulations. In *Stein's Method: Expository Lectures and Applications*. *IMS Lecture Notes Monogr. Ser.* **46**, 1–26. Beachwood, OH: Institute of Mathematical Statistics. MR2118600

Département de Mathématique
Université libre de Bruxelles
Campus Plaine
Boulevard du Triomphe, CP210
B-1050 Brussels
E-mail: chrisley@ulb.ac.be

Département de Mathématique
Université de Liège
Grande Traverse 12
Sart Tilman
B-4000 Liège
E-mail: yswan@ulg.ac.be