Parametric Stein operators and variance bounds

Christophe Ley* and Yvik Swan†

Université libre de Bruxelles and Université du Luxembourg

Abstract

Stein operators are differential 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. For specific choices of target distributions (including the Gaussian, Gamma and Poisson) we compare the operators hereby obtained with those provided by the classical approaches from the literature on Stein's method. 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.

AMS 2000 subject classifications: Primary 62E17; Secondary 60E15
Keywords: Chernoff inequality; Cramér-Rao inequality; parameter of interest; Stein characterization; Stein's method.

1 Introduction

Let g be a given target density (continuous or discrete) and let $X \sim g$. Choose d a probability metric (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 in the pathbreaking [37, 38]) advocates to first construct a suitable differential operator $f \mapsto \mathcal{T}(f, g)$ such that $X \sim g \iff \mathrm{E}[\mathcal{T}(f, g)(X)] = 0$ for all $f \in \mathcal{F}(g)$, with $\mathcal{F}(g)$ a specific (g-dependent) class of *test functions*, and then to use estimates on $\delta_g = \sup_{f\in\mathcal{F}(g)} |\mathrm{E}\left[\mathcal{T}(f, g)(W)\right]|$ (which, of course, is 0 if $W \sim g$) in order to estimate $d(W, X)$. Although mainly reserved to Gaussian approximation [4, 12, 31] and Poisson approximation [5], the method has also been proven in recent years to be very powerful for other types of approximation problems [10, 15, 17, 29, 30, 33–35].

The key to the success of the method lies in the properties of the operator $\mathcal{T}(\cdot, g)$ which, if the operator is well chosen, not only allow to obtain good estimates on the quantity δ_g but also guarantee that these in turn yield precise information on the probability distance $d(W, X)$.

*Département de Mathématique, Campus Plaine, Boulevard du Triomphe, CP210, B-1050 Brussels
†Unité de Recherche en Mathématiques, 6, rue Richard Coudenhove-Kalergi, L-1359 Luxembourg

There are several well-documented ways to construct a suitable Stein operator including the so-called *generator method* introduced by [3, 18] and the so-called *density approach* introduced in [38, 39]. For instance if $g = \phi$ the standard Gaussian density, then a routine application of the density approach gives the first-order operator $\mathcal{T}(f, \phi)(x) = -f'(x) + xf(x)$, while the generator approach brings the infinitesimal generator of the Ornstein-Uhlenbeck process on $\mathbb{R}$, that is, the second-order operator $\tilde{\mathcal{T}}(f, \phi)(x) = -f''(x) + xf'(x)$. If g is the rate-1 exponential distribution then suitable modifications of the density approach provide the operators $\mathcal{T}(f, g)(x) = -f'(x) + f(x)$ and $\bar{\mathcal{T}}(f, g)(x) = -xf'(x) + (x - 1)f(x)$; both have been used for exponential approximation problems [10, 33]. See also [15, 17, 27, 28, 36] for more examples and details.

Now consider a random variable $X \sim g$. Stein operators allow, in essence, to write general integration by parts formulas of the form

$$\mathrm{E}\left[f(X)h'(X)\right] = \mathrm{E}\left[\mathcal{T}(f, g)(X)h(X)\right] \tag{1.1}$$

which hold for all sufficiently regular test functions f and h. Setting $f = 1$ in (1.1) and applying the Cauchy-Schwarz inequality to the right-hand side we deduce that

$$\frac{(\mathrm{E}\left[h'(X)\right])^2}{\mathrm{E}\left[(\mathcal{T}(1, g)(X))^2\right]} \leq \mathrm{E}\left[(h(X))^2\right]$$

for all appropriate test functions h. This is a generalization of the celebrated Cramér-Rao inequality, with $\mathrm{E}\left[(\mathcal{T}(1, g)(X))^2\right]$ being some form of Fisher information for X. In particular if $g = \phi$ is the density of a standard Gaussian random variable and h has mean zero under ϕ, then $\mathcal{T}(1, \phi)(x) = -x$ and this last result particularizes to $(\mathrm{E}[h'(X)])^2 \leq \mathrm{Var}[h(X)]$. Chernoff [13, 14] used a method involving Hermite polynomials to prove that a converse inequality holds in the case of a Gaussian, namely

$$(\mathrm{E}[h'(X)])^2 \leq \mathrm{Var}[h(X)] \leq \mathrm{E}[(h'(X))^2] \tag{1.2}$$

with equality on both sides if and only if h is linear. Although Chernoff's technique of proof is tailored for a Gaussian target, an alternative proof hinging on the properties of the Gaussian Stein operator was obtained in [11] by Chen. Chen's approach was rapidly seen to be robust to a change in the distribution, and similar inequalities as (1.2) were shown to hold for many other distributions than the standard normal [7, 24]. The connection between the so-called Stein type identities and extensions of the bound (1.2) to arbitrary target distributions has now been explored in quite some detail [6–9, 11, 14, 32].

Remark 1.1. *In the case of functionals of a Gaussian, the upper and lower bounds in* (1.2) *were shown in [20] – again through an argument specifically tailored for a Gaussian target – to be the first terms in an infinite series expansion. Recently [1] extended the latter series expansion for the lower bound from a Gaussian target to the class of Pearson distributions (and Ord distributions in the discrete case) by studying the properties of the Stein operators for distributions satisfying Pearson's (or Ord's) definition.*

Remark 1.2. *Bounds such as* (1.2) *and its variations for alternative targets have deep theoretical and practical implications. These are connected to the classical isoperimetric problem and thus also to logarithmic Sobolev inequalities [2]. Tight estimates on the constants in these inequalities (called Poincaré constants) provide crucial quantitative information on the properties of the distribution of the random variable.*

In this 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. More precisely, given a target g and a parameter θ we identify a maximal class $\mathcal{F}(g;\theta)$ of test functions (maximal because the conditions are minimal) and a differential operator $\mathcal{T}_\theta(\cdot, g)$ such that

$$X \sim g(\cdot;\theta) \iff \mathrm{E}\left[\mathcal{T}_\theta(f,g)(X)\right] = 0$$

for all $f \in \mathcal{F}(g;\theta)$. 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 (1.2) to a wide variety of target distributions g. One of the strong points of our bounds is that, when applied to specific distributions such as the exponential, Gamma or Gaussian, the results obtained via our technique are systematically competitive with (and sometimes improve on) the best bounds available in the literature. 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 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

$$\int g(x;\theta)d\mu(x) = 1 \tag{2.1}$$

for all $\theta \in \Theta$. 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 $\theta \mapsto f(\cdot;\theta)g(\cdot;\theta)$ is differentiable in the sense of distributions over Θ_0;

(iii) there exists an integrable function $h : \mathbb{R} \to \mathbb{R}^+$ such that for all $\theta \in \Theta_0$ we have $|\partial_\theta(f(x;\theta)g(x;\theta))| \leq h(x)$ 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* $g(\cdot;\theta_0)$.

The conditions imposed in Definition 2.1 are, in a sense, too stringent, and minimal conditions on the test functions f are obtained simply by requiring that one can take derivatives (with respect to θ) under the integral sign. Conditions (ii) and (iii) are natural sufficient assumptions for this manipulation to be allowed; see e.g. [25].

We now state the main result of this section. The proof 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 *and if* $\mathrm{E}[\mathcal{T}_{\theta_0}(f,g)(X)] = 0$ *for all* $f \in \mathcal{F}(g;\theta_0)$, *then* $X \,|\, X \in S \stackrel{\mathcal{D}}{=} Z_{\theta_0}$.

In the next sections we consider three well-known types of parameters, namely location, scale and skewness (in each case for absolutely continuous target distributions), and use Theorem 2.1 to construct a selection of relevant and tractable Stein operators which are, in a sense, natural with respect to the choice of parameter. 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 $\mu_0 \in \mathbb{R}$ (typically one takes $\mu_0 = 0$) and consider densities of the form

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

for g_0 some positive function integrating to 1 over its support. We denote by $\mathcal{G}_{\mathrm{loc}}$ the collection of g_0's for which μ-parametric densities of the form (2.2) belong to $\mathcal{G}(\mathbb{R}, \mu_0)$.

Clearly, in the present context, Condition (i) of Definition 2.1 holds most naturally for test functions of the form $f(x;\mu) = f_0(x-\mu)$ for which we also have

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

for all $(x,\mu) \in \mathbb{R} \times \mathbb{R}$ (we write ∂_x and ∂_μ the weak derivatives with respect to x and μ, respectively). Let $g_0 \in \mathcal{G}_{\mathrm{loc}}$. We then define $\mathcal{F}_{\mathrm{loc}}(g_0;\mu_0)$ the collection of all $f_0 : \mathbb{R} \to \mathbb{R}$ such that (i) $\int_{\mathbb{R}} f_0(x-\mu)g_0(x-\mu)dx = \int_{\mathbb{R}} f_0(x)g_0(x)dx = c_{f_0}$ 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 $\left|\partial_y (f_0(y-\mu)g_0(y-\mu))|_{y=x}\right| \leq h(x)$ over $\mathbb{R}$ for all $\mu \in \Theta_0$, some bounded neighborhood of μ_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*

$$\mathcal{T}_{\mu_0;\mathrm{loc}}(f_0, g_0) : \mathbb{R} \to \mathbb{R} : x \mapsto \frac{-\partial_y (f_0(y-\mu_0) g_0(y-\mu_0))|_{y=x}}{g_0(x-\mu_0)}, \tag{2.4}$$

for $f_0 \in \mathcal{F}_{\mathrm{loc}}(g_0; \mu_0)$ and with ∂_y the derivative in the sense of distributions with respect to y.

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 $\mu_0 = 0$ 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 $H_0(x) = 1$, $H_{n+1}(x) = -H_n'(x) + x H_n(x)$ (that is, $H_n(x)$ is the nth Hermite polynomial) and consider functions of the form $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R} : (x, \mu) \mapsto f(x; \mu) := H_n(x-\mu) f_0(x-\mu)$, where $f_0 : \mathbb{R} \to \mathbb{R}$ is chosen such that $f \in \mathcal{F}_{\mathrm{loc}}(\phi; 0)$. Restricting the operator $\mathcal{T}_{\mathrm{loc}}(\phi, \cdot)$ to this collection of f's, we find

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

This family of operators was discovered by [16].

Next, 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 $\mu_0 = 0$ we get the operator

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

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 [39] and used in [10] under the restriction $f_0(0) = 0$. More generally, when g belongs to the (continuous) *exponential family* (see [25] for a precise definition), these location-based manipulations allow to retrieve the known operators discussed e.g. in [21], [22] or [25].

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

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

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 most naturally for test functions of the form $f(x; \sigma) = f_0(\sigma x)$ for which we have the relationship

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

for all $(x, \sigma) \in \mathbb{R} \times \mathbb{R}_0^+$. Let $g_0 \in \mathcal{G}_{\mathrm{sca}}$. We then define $\mathcal{F}_{\mathrm{sca}}(g_0; \sigma_0)$ 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 $\left| \partial_y (y f_0(\sigma y) g_0(\sigma y))|_{y=x} \right| \leq h(x)$ 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*

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

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.

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

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

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

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

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

after setting $\sigma_0 = 1$. This scale-based operator has first been exploited in [10]. More generally, choosing g the probability density function (pdf) of a Gamma distribution with shape $a > 0$ we obtain

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

a variant of the Gamma operator used, e.g., by [30].

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 [23]. These laws are a particular case of the construction given in [26] 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

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

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 LP-densities the skewed distributions (2.9).

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 $\mathcal{G}_{\mathrm{skew}}(H_\delta)$ 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 $f(x;\delta) = f_0(H_\delta(x))$. Let $g_0 \in \mathcal{G}_{\text{skew}}(H_\delta)$. We then define $\mathcal{F}_{\text{skew}}(g_0; H_{\delta_0})$ the collection of all $f_0 : \mathbb{R} \to \mathbb{R}$ such that (i) $\int_{\mathbb{R}} f_0(H_\delta(x))(H_\delta)'(x)g_0(H_\delta(x))dx = \int_{\mathbb{R}} f_0(x)g_0(x)dx = c_{f_0}$ 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 [23]) 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

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

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

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

Further specifying $g_0 = \phi$ the standard Gaussian pdf and taking $f_0(x) = \sqrt{1+x^2}f_1(x)$ with f_1 some suitable function 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 $S = [N] := \{0,\ldots,N\}$ for some $N \in \mathbb{N}_0 \cup \{\infty\}$ not depending on θ and that the function $\theta \mapsto g(x;\theta)$ is weakly differentiable around θ_0 at all $x \in [N]$.

Condition (i) of Definition 2.1 here holds for test functions of the form

$$f(x;\theta) = \frac{D_x^+\left(f_0(x)\dfrac{g(x;\theta)}{g(0;\theta)}\right)}{g(x;\theta)} \tag{2.11}$$

(with $D_x^+(f(x)) = f(x+1) - f(x)$ the forward difference operator). The combination of D_x^+ and ∂_θ permits us to exchange derivatives with respect to the variable and the parameter, that is, for f of the form (2.11) we have the relation

$$\partial_\theta(f(x;\theta)g(x;\theta)) = D_x^+(f_0(x)\partial_\theta(g(x;\theta)/g(0;\theta)))$$

for all $(x,\theta) \in [N] \times \mathbb{R}$. Let $g \in \mathcal{G}_{\mathrm{dis}}$. We then define $\mathcal{F}_{\mathrm{dis}}(g;\theta_0)$ 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*

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

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

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 $\partial_\lambda\big(g(x;\lambda)/g(0;\lambda)\big)\big|_{\lambda=\lambda_0} = \lambda_0^{x-1}/(x-1)!$ so that

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

which is (up to the scaling factor) the usual operator for the Poisson. Setting $g(x;p) = (1-p)^x p\,\mathbb{I}_{\mathbb{N}}(x)$ the geometric $Geom(p)$ distribution, we get

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

Finally, for the binomial $Bin(n,p)$, we obtain the p-characterizing operator

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

These last two operators are not new, and can be obtained (up to scaling factors) via the generator approach [19].

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 $S = \{a, a+1, \ldots, b-1, b\}$.)

We single out the subclass $\mathcal{F}_1(g; \theta_0) \subset \mathcal{F}(g; \theta_0)$ (often written $\mathcal{F}_1$, whenever no ambiguity ensues) 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

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

satisfies the boundary conditions

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

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

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

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^+.

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 $X^\star_{f,\theta_0} \sim g^\star(\cdot; \theta_0) = f(\cdot; \theta_0) g(\cdot; \theta_0)$. We write $\varphi_{\theta_0, g^\star}(x) := \partial_\theta (\log(g^\star(x; \theta)))\big|_{\theta=\theta_0} (= \mathcal{T}_{\theta_0}(f, g)(x)/f(x; \theta_0))$ 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*

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

for all $h \in C_0^1(\mathbb{R})$. If, furthermore, $x \mapsto \varphi_{\theta_0, g^\star}(x)$ is strictly monotone and strongly differentiable over its support then

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

for all $h \in C_0^1(\mathbb{R})$. 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 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 $\tilde{f}(x;\theta_0) \le 0$ (resp., $\tilde{f}(x;\theta_0) \ge 0$) 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 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. This is easily seen in the case of the rate-1 exponential distribution with location parameter μ. There the dependence of the support on the parameter implies the appearance of a Dirac delta in the expression of the location operator (2.6); as a consequence Theorem 3.1 does not apply to this particular case. We contend that this breakdown is not a drawback of our approach but rather one of its strengths and we will see that, despite this restriction, the bounds we obtain are as good as if not better than those already available in the literature (see the discussion at the end of this section).

In practice, the problem is avoided by assuming that the support of $g(\cdot;\theta)$ is either open or does not depend on θ. Then $f(x;\theta) = 1$ 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.

Proposition 3.1 (Location, scale and skewness variance bounds)**.** *Consider a θ-parametric density $g \in \mathcal{G}(\mathbb{R}, \theta_0)$ and let $X \sim g(\cdot;\theta_0)$.*

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 $g_0 \in C^1_0(S)$ with open support S. Then the exchanging function for $f(x;\mu) = f_0(x-\mu) \in \mathcal{F}_{\mathrm{loc}}(g_0;\mu_0)$ around μ is $\tilde{f}(x;\mu) = -f_0(x-\mu)$. The location-score function (expressed in terms of $y = x-\mu$) is*

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

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

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

for $h \in C^1_0(\mathbb{R})$, with $\mathcal{I}_{\mathrm{loc}}(g_0) := \mathrm{E}\left[(\varphi_{g_0,\mathrm{loc}}(X-\mu_0))^2\right]$.

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 $g_0 \in C^1_0(S)$ with either open support S or support S invariant under scale change. Then the exchanging function for $f(x;\sigma) = f_0(\sigma x) \in \mathcal{F}_{\mathrm{sca}}(g_0;\sigma_0)$ 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*

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

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

$$\frac{(\mathrm{E}\left[h'(X)X\right])^2}{\sigma_0^2\mathcal{I}_{\text{sca}}(g_0)} \le \mathrm{Var}\left[h(X)\right] \le \mathrm{E}\left[\frac{(h'(X))^2X}{-\sigma_0^2\varphi'_{g_0,\text{scale}}(\sigma_0X)}\right] \tag{3.7}$$

for $h \in C_0^1(\mathbb{R})$, with $\mathcal{I}_{\text{sca}}(g_0) := \mathrm{E}\left[(\varphi_{g_0,\text{scale}}(\sigma_0X))^2\right]$.

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 $f(x;\sigma) = f_0(S_\delta(x)) \in \mathcal{F}_{\text{skew}}(g_0; S_{\delta_0})$ 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*

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

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

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

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

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 Proposition 3.1 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 $\mathbb{R}$).

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 Proposition 3.1 for $\mu_0 = 0$ (location case), $\sigma_0 = \sigma$ (scale case) and $\delta_0 = 0$ (skewness case) we get

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

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

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

We thus sequentially obtain the location-based variance bounds

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

with equality if and only if h is linear (this is the well-known bound (1.2); 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

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

with equality if and only if $h(x) \propto 1 - x^2$ (this bound is given in [24]) and also the skewness-based bound

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

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

Next 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 $x > 0$). Thus, by Proposition 3.1 for $\sigma_0 = \lambda$ we get

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

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

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

For the sake of comparison, [7] proposes the lower and upper bounds

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

while [24] proposes

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

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

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

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).

Finally take $g_0(x) = \frac{1}{\Gamma(a)}x^{a-1}e^{-x}\mathbb{I}_{[0,\infty)}(x)$ the pdf 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 $a \leq 1$. 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 Proposition 3.1 under the respective restrictions on a and for $\mu_0 = 0$ (location case) and $\sigma_0 = b$ (scale case), we get

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

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

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

This yields the following : location-based bounds

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

and scale-based bounds

$$\frac{1}{a}\left(\mathrm{E}\left[Xh'(X)\right]\right)^2 \leq \mathrm{Var}\left[h(X)\right] \leq \frac{1}{b}\mathrm{E}\left[X(h'(X))^2\right]. \tag{3.13}$$

On the one hand [7] only proposes a lower bound (which concurs with ours). On the other hand, [24] proposes for $a > 2$

$$\max\left(\frac{a-2}{b^2}\left(\mathrm{E}\left[h'(X)\right]\right)^2, \frac{1}{a}\left(\mathrm{E}\left[Xh'(X)\right]\right)^2\right) \leq \mathrm{Var}\left[h(X)\right] \leq \frac{1}{b}\mathrm{E}\left[X(h'(X))^2\right]. \tag{3.14}$$

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

$$\mathrm{Var}(h(X)) \leq d\,\mathrm{E}\left[(h'(X))^2\right] \tag{3.15}$$

for some positive constant d. If the special case $f = 1$ is admissible then, trivially, $d = d_{g,\theta_0} = \sup_{x \in S}(-\tilde{f}(x;\theta_0)/\varphi'_{\theta_0,g^\star}(x))$ 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 3.2. *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 $\epsilon > 0$ such that $-(\log g(x))'' \geq \epsilon > 0$ then (3.15) holds with $d_{g,\mu_0} = \frac{1}{\epsilon}$.*

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

$$\frac{\tilde{f}(x;\mu_0)}{-\varphi'_{\mu_0,g^\star}(x)} = \frac{1}{-\frac{g''(x-\mu_0)}{g(x-\mu_0)} + \left(\frac{g'(x-\mu_0)}{g(x-\mu_0)}\right)^2} = \frac{1}{-(\log g(x-\mu_0))''}.$$

The conclusion follows from (3.15). □

Note that the assumptions of Proposition 3.2 hold if $g(x) = e^{-\psi(x)}$ for $\psi(x)$ a strict convex function, i.e. if g is strongly unimodal on $\mathbb{R}$. We hereby recover Lemma 2.1 from [24]. 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 $\epsilon = 1/\sigma^2$ and we re-obtain the well-known upper bound $\mathrm{Var}\,(h(X)) \leq \sigma^2\mathrm{E}\left[(h'(X))^2\right]$.

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

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

The boundary condition (3.2) therefore allows 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 $X^\star_{f,\theta_0} \sim g^\star(\cdot;\theta_0) = f(\cdot;\theta_0)g(\cdot;\theta_0)$. We write $\varphi_{\theta_0,g^\star}(x) := \partial_\theta(\log(g^\star(x;\theta)))\big|_{\theta=\theta_0}(= \mathcal{T}_{\theta_0}(f,g)(x)/f(x;\theta_0))$ 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*

$$\mathrm{Var}\left[h(X^\star_{f,\theta_0})\right] \geq \frac{\left(\mathrm{E}\left[D_x^+(h(x))|_{x=X}\tilde{f}(X;\theta_0)\right]\right)^2}{\mathcal{I}(\theta_0, g^\star)} \tag{3.16}$$

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

Take $g(x;\lambda) = e^{-\lambda}\lambda^x/x!\mathbb{I}_{\mathbb{N}}(x)$ the pdf of the Poisson distribution. Then we have $\partial_\lambda g(x;\lambda) = -D_x^+\left(\frac{x}{\lambda}g(x;\lambda)\right)$; 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

$$\mathrm{Var}\left[h(X)\right] \geq \frac{1}{\lambda}\mathrm{E}\left[XD_x^+(h(x))|_{x=X}\right]^2,$$

with equality if and only if $h(x) \propto -1 + x/\lambda$ on $\mathbb{N}$. This last bound is, to the best of our knowledge, new.

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,

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

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

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

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

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

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

$$\begin{aligned}\int_S l_A(x;u)g(x;u)dx &= \mathrm{P}(Z_u \in A \cap S) - \mathrm{P}\left(Z_u \in A\right)\mathrm{P}(Z_u \in S)\\ &= 0.\end{aligned}$$

Hence f_A satisfies Condition (i). Condition (ii) is easily checked. Regarding Condition (iii), one sees that $\partial_t\left(f_A(x;t)g(x;t)\right)|_{t=\theta} = l_A(x;\theta)g(x;\theta)$. 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. $\square$

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

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

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

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

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

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

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,

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

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

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

From (3.3) we then get

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

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

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

which is the claim. □

ACKNOWLEDGEMENTS

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.

References

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

[2] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.

[3] A. D. Barbour. Stein's method for diffusion approximations. *Probab. Theory Related Fields*, 84(3):297–322, 1990.

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

[5] A. D. Barbour, L. Holst, and S. Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability.* The Clarendon Press Oxford University Press, New York, 1992. Oxford Science Publications.

[6] A. Borovkov and S. Utev. On an inequality and a related characterization of the normal distribution. *Theory of Probability & Its Applications*, 28(2):219–228, 1984.

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

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

[9] T. Cacoullos, V. Papathanasiou, and S. A. Utev. Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.*, 22(3):1607–1618, 1994.

[10] S. Chatterjee, J. Fulman, and A. Roellin. Exponential approximation by exchangeable pairs and spectral graph theory. *ALEA*, 8:1–27, 2011.

[11] L. H. Y. Chen. An inequality for multivariate normal distribution. Technical report, MIT, 1980.

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

[13] H. Chernoff. The identification of an element of a large population in the presence of noise. *Ann. Statist.*, 8(6):1179–1197, 1980.

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

[15] C. Döbler. Stein's method of exchangeable pairs for absolutely continuous, univariate distributions with applications to the polya urn model. Preprint, arXiv:1207.0533, 2012.

[16] L. Goldstein and G. Reinert. Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.*, 18(1):237–260, 2005.

[17] L. Goldstein and G. Reinert. Stein's method and the beta distribution. Preprint, arxiv:1207.1460, 2012.

[18] F. Götze. On the rate of convergence in the multivariate clt. *Ann. Probab.*, 19(2):724–739, 1991.

[19] S. Holmes. Stein's method for birth and death chains. In *Stein's method: expository lectures and applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 45–67. Inst. Math. Statist., Beachwood, OH, 2004.

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

[21] H. M. Hudson. A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.*, 6:473–484, 1978.

[22] J. T. Hwang. Improving upon standard estimators in discrete exponential families with applications to poisson and negative binomial cases. *Ann. Statist.*, 10:857–867, 1982.

[23] M. C. Jones and A. Pewsey. Sinh-arcsinh distributions. *Biometrika*, 96:761–780, 2009.

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

[25] E. L. Lehmann and G. Casella. *Theory of point estimation*, volume 31. Springer-Verlag, New York, 1998.

[26] C. Ley and D. Paindaveine. Multivariate skewing mechanisms: a unified perspective based on the transformation approach. *Statist. Probab. Lett.*, 80:1685–1694, 2010.

[27] C. Ley and Y. Swan. Stein's density approach for discrete distributions with applications to information inequalities. Preprint, arXiv:1211.3668, 2012.

[28] C. Ley and Y. Swan. Stein's density approach and information inequalities. *Electron. Comm. Probab.*, 18(7):1–14, 2013.

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

[30] I. Nourdin and G. Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 145(1-2):75–118, 2009.

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

[32] N. Papadatos and V. Papathanasiou. Unified variance bounds and a stein-type identity. *Probability and Statistical Models with Applications*, pages 87–100, 2001.

[33] E. Peköz and A. Roellin. New rates for exponential approximation and the theorems of rényi and yaglom. *Ann. Probab.*, 39(2):587 – 608, 2011.

[34] E. Peköz, A. Roellin, and N. Ross. Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Probab.*, 2012, to appear.

[35] A. Picket. *Rates of convergence of χ^2 approximations via Stein's method.* PhD thesis, Lincoln College, University of Oxford, 2004.

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

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

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

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