Some simple variance bounds from Stein's method

Some simple variance bounds from Stein's method

Fraser Daly, Fatemeh Ghaderinezhad, Christophe Ley and Yvik Swan

Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh EH14 4AS, UK

Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Krijgslaan 281, S9, Campus Sterre, 9000 Gent, Belgium

Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Krijgslaan 281, S9, Campus Sterre, 9000 Gent, Belgium

Department of Mathematics, Université libre de Bruxelles, Boulevard du Triomphe, CP210, B-1050 Bruxelles
E-mail address: `f.daly@hw.ac.uk, fatemeh.ghaderinezhad@ugent.be, christophe.ley@ugent.be, yvik.swan@ulb.ac.be`

Abstract: Using coupling techniques based on Stein's method for probability approximation, we revisit classical variance bounding inequalities of Chernoff, Cacoullos, Chen and Klaassen. Our bounds are immediate in any context wherein a Stein identity is available. After providing illustrative examples for a Gaussian and a Gumbel target distribution, our main contributions are new variance bounds in settings where the underlying density function is unknown or intractable. Applications include bounds for analysis of the posterior in Bayesian statistics, bounds for asymptotically Gaussian random variables using zero-biased couplings, and bounds for random variables which are New Better (Worse) than Used in Expectation.

1. Introduction

Poincaré (or isoperimetric) inequalities, giving upper bounds on the variance of a function of a random variable, have a long and rich history, beginning with the work of Chernoff [8]. Chernoff proved that if X has a centred Gaussian distribution with variance σ^2, then

$$\mathrm{Var}[g(X)] \leq \sigma^2 \mathbb{E}[(g'(X))^2], \tag{1.1}$$

for any absolutely continuous function $g : \mathbb{R} \to \mathbb{R}$ such that $g(X)$ has finite variance. This inequality has since been generalized by many authors, including Cacoullos [3], Chen [5] and Klaassen [13]. To accompany these upper variance bounds, many of these authors have also established corresponding lower bounds, in the form of generalized Cramér-Rao inequalities. In particular, in the centred Gaussian case we have

$$\mathrm{Var}[g(X)] \geq \sigma^2 \mathbb{E}[g'(X)]^2 \tag{1.2}$$

for all absolutely continuous g such that $\mathrm{Var}[g(X)] < \infty$ and $\mathbb{E}[|g'(X)|] < \infty$; see [3, Proposition 3.1]. The above cited works represent early entries in what is now a vast literature; we refer to [9, 10] for recent overviews of this large body of work.

The purpose of the present article is to revisit these classical variance bounding inequalities in light of the coupling techniques at the heart of Stein's method for probability approximation. We refer the reader unfamiliar with Stein's method to [6, 17] and [14] for recent introductions to this area. Stein's method is a suite of techniques which have principally been exploited to derive explicit error bounds for probabilistic approximations, classically in the Gaussian and Poisson settings and more recently across a wide variety of univariate, multivariate and process-level limiting objects. One fruitful approach to applying Stein's technique involves the construction of couplings closely related to the appropriate limiting distribution. For example, the zero-biased coupling introduced by [11] has the mean-zero Gaussian distribution as

a fixed point and has been used to derive explicit bounds in Gaussian approximation in a variety of applications; many applications of zero-biased (and other) couplings in Gaussian approximation using Stein's method are detailed in [6].

By exploiting such coupling constructions, we derive upper and lower variance bounds, including *weighted* Poincaré inequalities (see, e.g., the upper bound (1.6) below, which has a weight ω_γ, compared to the unweighted case (1.1)). These variance bounds may be applied in a wide variety of settings, including many in which the density of the underlying random variable is unknown or intractable. Our key coupling identity is (1.3) below, from which our variance bounds will follow. The identity (1.3) generalises many of the couplings (including zero-biasing) used in Stein's method, and therefore allows us to treat a wide variety of applications and examples. Employing the zero-biased coupling will allow us to establish explicit variance bounds for a range of situations in which the underlying random variable is known to be asymptotically Gaussian, but we emphasise that the more general coupling defined in (1.3) will allow us to treat many examples unrelated to the Gaussian setting.

In Sections 2–4 below we will consider a variety of settings in which variance bounds can be derived by exploiting our coupling identity (1.3). Before doing so, we use the remainder of this section to state this identity, and demonstrate how it can be used to establish upper and lower variance bounds in the spirit of Chernoff, Cacoullos, Chen and Klaassen.

Let W be a real-valued random variable on some fixed probability space and let $L^r(W), r \geq 1$, be the collection of real-valued functions γ such that $\mathbb{E}|\gamma(W)|^r < \infty$. Let $\gamma \in L^1(W)$ with $\mathbb{E}[\gamma(W)] = 0$. We say that a pair of random variables (T_1, T_2) (living on the same probability space as W) form a Stein coupling for W with respect to γ if

$$\mathbb{E}\left[\gamma(W)\phi(W)\right] = \mathbb{E}\left[T_1\phi'(T_2)\right] \tag{1.3}$$

for all differentiable test functions $\phi \in C$ with $C \subset C^1(\mathbb{R})$ some appropriately chosen class of functions. Although the choice $C = C_0^\infty(\mathbb{R})$ is always allowed, it will generally be interesting to use C as wide as possible – in particular one would wish that polynomials x^k belong to C for some k. In order to keep the text light we will follow the tradition in the literature wherein one rather makes use of the expression "where C is the class of functions for which expectations exist". We note that in using this expression we are also including assumptions (such as differentiability of ϕ) necessary for existence of any expressions within the expectations. We also note here that what we have called a Stein coupling is not the same as the Stein couplings introduced by [7].

We begin by showing an elementary argument allowing us to use (1.3) to obtain tight upper variance bounds. To this end, suppose that γ is a strictly increasing and differentiable function. Then in particular it is invertible and $\gamma^{-1}(0)$ is well-defined. Let g be a real-valued differentiable function such that $\mathrm{Var}[g(W)]$ is finite. Following [15] we write

$$\mathrm{Var}[g(W)] \leq \mathbb{E}\left[\left(g(W) - g(\gamma^{-1}(0))\right)^2\right] = \mathbb{E}\left[\left(\int_0^{\gamma(W)} \frac{g'(\gamma^{-1}(u))}{\gamma'(\gamma^{-1}(u))}\mathrm{d}u\right)^2\right]$$

$$\leq \mathbb{E}\left[\gamma(W)\int_0^{\gamma(W)} \left(\frac{g'(\gamma^{-1}(u))}{\gamma'(\gamma^{-1}(u))}\right)^2 \mathrm{d}u\right],$$

where the equality follows by differentiability of g and the subsequent inequality via the Cauchy-Schwarz inequality. Applying (1.3) for differentiating integrals we deduce the general upper variance bound

$$\mathrm{Var}[g(W)] \leq \mathbb{E}\left[\frac{T_1}{\gamma'(T_2)}\left(g'(T_2)\right)^2\right], \tag{1.4}$$

which holds as soon as the function $x \mapsto \int_0^{\gamma(x)} \left(g'(\gamma^{-1}(u))/\gamma'(\gamma^{-1}(u))\right)^2 \mathrm{d}u$ belongs to the class C discussed following (1.3). Note that inequality (1.4) also holds if in (1.3) we replace the equality sign by an increasing inequality.

Identity (1.3) can also readily be combined with the Cauchy-Schwarz inequality to obtain lower variance bounds. To this end, consider a mean zero function γ, for which we have

$$\left(\mathbb{E}\left[\gamma(W)g(W)\right]\right)^2 = \left(\mathbb{E}\left[\gamma(W)(g(W) - \mathbb{E}[g(W)])\right]\right)^2 \leq \mathbb{E}\left[\gamma(W)^2\right]\mathrm{Var}[g(W)].$$

Then from (1.3) we deduce

$$\mathrm{Var}[g(W)] \geq \frac{(\mathbb{E}\left[T_1 g'(T_2)\right])^2}{\mathrm{Var}\left[\gamma(W)\right]} \tag{1.5}$$

for all $g \in C$. As above, we note that inequality (1.5) also holds if in (1.3) we replace the equality sign by a decreasing inequality.

One of the advantages of bounds such as (1.4) and (1.5) is their ease of use, which springs from the many "Stein identities" (1.3) in the literature. Before tackling more challenging situations in the coming sections, we demonstrate the bounds we obtain in two specific instances, beginning with the Gaussian setting. These examples are used primarily as illustrations of the bounds that can be obtained from (1.4) and (1.5) in some concrete cases, and to draw connections between our inequalities and others available in the literature.

Example 1.1 (Gaussian upper bounds). *Let $X \sim \mathcal{N}(0,1)$, and define for each $f \in L^1(X)$ the operator $\mathcal{L}f(x) = e^{x^2/2} \int_{-\infty}^{x} (f(u) - \mathbb{E}[f(X)])e^{-u^2/2}\mathrm{d}u$. A simple adaptation of the proof of [9, Proposition 2.29] shows that the operator $\mathcal{L}$ satisfies*

$$\mathbb{E}\left[\gamma(X)\phi(X)\right] = \mathbb{E}\left[\left(-\mathcal{L}\gamma(X)\right)\phi'(X)\right]$$

for all centered monotone $\gamma \in L^2(X)$ and all absolutely continuous test functions $\phi \in L^2(X)$ such that $\phi' \in L^1(\gamma)$. We collect all these functions into the class C discussed below (1.3). We can therefore apply (1.4) with $T_1 = -\mathcal{L}\gamma(X)$ and $T_2 = X$, leading to the (weighted Poincaré) inequality

$$\mathrm{Var}[g(X)] \leq \mathbb{E}\left[\frac{-\mathcal{L}\gamma(X)}{\gamma'(X)}(g'(X))^2\right] =: \mathbb{E}[\omega_\gamma(X)(g'(X))^2] \tag{1.6}$$

which holds for all centred, increasing and differentiable functions γ belonging to $L^2(X)$ and all g satisfying the condition given below (1.4) (it may be of interest to clarify this class using e.g. the results from [9, Section 2.3] but, in the interest of brevity, we simply remark that all differentiable functions with compact support belong to this class – larger classes of functions can be attained on a case-by-case basis). Obviously, different choices of γ in (1.6) lead to different inequalities. For example, taking $\gamma(x) = x$ leads to $\omega_\gamma(x) = 1$, reproducing (1.1), as expected, whereas $\gamma(x) = x^3$ leads to $\omega_\gamma(x) = \frac{1}{3}\left(1 + \frac{2}{x^2}\right)$. Following [1], we can also consider $\gamma(x) = xe^{bx^2/2}$, which leads to a weight $\omega_\gamma(x) = ((b-1)(1+bx^2))^{-1}$, for $b < 1$. We refer the interested reader to Section 2.2 of [1] for a more detailed discussion of weighted Poincaré inequalities with weights of this form, and in particular for a discussion of optimality of such weights. See also [21] and [9] for more detail, as well as [22] for other identities such as (1.3). Finally, we mention that lower bounds can also be readily obtained from (1.5), for instance with the help of the identities available in [12, Section 4.1]; we do not detail this here.

Example 1.2 (Gumbel bounds). *Let W be a standard Gumbel random variable with CDF $P(x) = e^{-e^{-x}}$ and PDF $p(x) = e^{-e^{-x}-x}$ on $\mathbb{R}$. Then $\mathbb{E}[(1 - e^{-W})\phi(W)] = \mathbb{E}[\phi'(W)]$ for all functions $\phi \in C$ the collection of differentiable functions such that $\phi' \in L^1(W)$; see [14, Section 6.5]. We can therefore apply (1.4) and (1.5) with $T_1 = 1$, $T_2 = W$ and $\gamma(x) = 1 - e^{-x}$ leading to the inequalities*

$$\mathbb{E}[g'(W)]^2 \leq \mathrm{Var}[g(W)] \leq \mathbb{E}\left[e^W(g'(W))^2\right]$$

for all appropriate g. An analogous identity holds for the random variable $X_n = \max(\xi_1, \ldots, \xi_n) - \log(n+1)$ with $(\xi_i)_{i \geq 1}$ an IID sequence of rate 1 exponential random variables. Indeed, the same arguments from [14, Section 6.5] show that $\mathbb{E}[(1 - e^{-X_n})\phi(X_n)] = \mathbb{E}[(1 - \frac{e^{-X_n}}{n+1})\phi'(X_n)]$ for all appropriate ϕ, so that we may now apply (1.4) and (1.5) with $T_1 = \frac{e^{-X_n}}{n+1} - 1$, $T_2 = X_n$ and $\gamma(x) = 1 - e^{-x}$. Upon noting that $\mathrm{Var}[e^{-X_n} - 1] = \frac{n}{n+2}$ we obtain the bounds

$$\frac{n+2}{n}\mathbb{E}\left[\left(1 - \frac{e^{-X_n}}{n+1}\right)g'(X_n)\right]^2 \leq \mathrm{Var}[g(X_n)] \leq \mathbb{E}\left[e^{X_n}\left(1 - \frac{e^{-X_n}}{n+1}\right)(g'(X_n))^2\right].$$

The similarity between the upper and lower bounds for W and X_n comes as no surprise: it is a well-known fact that X_n converges in distribution to W.

One could of course obtain many more inequalities from Stein identities; it may also be of interest to study in more detail the class of functions on which the different inequalities hold. We shall not do this here and the rest of this paper is devoted to proposing less direct applications. In Section 2 we use the framework of Stein kernels to express suitable couplings and obtain a family of inequalities directly inspired by a classical result from Cacoullos [3]. Our main application is towards densities arising in a Bayesian framework. Section 3 makes use of zero-biased couplings to derive variance bounds suitable for random variables which are asymptotically Gaussian. Here, as in the next section, explicit knowledge of the density is not required for bounds to be obtained. Finally, in Section 4 we consider random variables satisfying certain stochastic or convex ordering assumptions. Some proofs and additional examples illustrating the results of Section 2 are deferred to the appendices.

2. The Stein kernel and a bound of Cacoullos

Suppose that the target random variable W has a differentiable density p with connected support (this excludes examples such as the Laplace distribution) and let P be the corresponding CDF. Following, for example, [4] and [9], we define the *Stein kernel* of W as the function τ satisfying

$$\text{Cov}\left[W, \phi(W)\right] = \mathbb{E}\left[\tau(W)\phi'(W)\right] \tag{2.1}$$

for all absolutely continuous functions ϕ such that either integral is defined; this classical object in the theory of Stein's method has been studied in detail in the literature, and we refer to [21, 9] for extensive discussions. In the notation of Section 1, equation (2.1) implies that we can take $\gamma(x) = x - \mathbb{E}[W]$, $T_1 = \tau(W)$ and $T_2 = W$ in (1.3). Note that $\mathbb{E}[\tau(W)] = \text{Var}[W]$. Applying (1.4) and (1.5) we get that

$$\frac{\mathbb{E}\left[\tau(W)g'(W)\right]^2}{\text{Var}\left[W\right]} \leq \text{Var}[g(W)] \leq \mathbb{E}\left[\tau(W)\left(g'(W)\right)^2\right], \tag{2.2}$$

and this can be seen to hold for all absolutely continuous g such that $g(X)$ has finite variance and $|g'(X)|$ has finite mean (see e.g. [21]). This is nothing but a restatement of classical bounds already available in [3]. Since the Stein kernel is a positive function, we can re-express (2.2) in the form

$$\frac{\mathbb{E}_\tau\left[g'(W)\right]^2}{\text{Var}\left[W\right]} \leq \text{Var}[g(W)] \leq \mathbb{E}_\tau\left[\left(g'(W)\right)^2\right],$$

where $\mathbb{E}_\tau$ is expectation under the new measure $\tau(\cdot)dP(\cdot)$.

Of course, for (2.2) to be of use it remains to identify situations in which the Stein kernel has an agreeable form, hereby resulting in informative bounds. In the following three examples we propose several such instances.

Example 2.1. *Following [18], it is easy to see that if $W = n^{-1/2}\sum_{i=1}^n X_i$, where the X_i are centred, independent random variables with Stein kernel $\tau_i(\cdot)$ and common variance σ^2, then $\tau_W(w) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\tau_i(X_i)\,|\,W = w]$ is a Stein kernel for W. Specialising to the the IID case, if the X_i are copies of X_1 with kernel $\tau_1(\cdot)$, (2.2) becomes*

$$\frac{\mathbb{E}\left[\tau_1(X_1)g'(W)\right]^2}{\sigma^2} \leq \text{Var}[g(W)] \leq \mathbb{E}\left[\tau_1(X_1)(g'(W))^2\right]$$

for all absolutely continuous g such that all expectations exist. We now aim to interpret these upper and lower bounds by comparison with the Gaussian upper and lower bounds from (1.1) and (1.2): this is natural because, by the CLT, the law of W is approximately Gaussian for n sufficiently large.

First the upper bound. We recall that $\mathbb{E}[\tau_1(X_1)] = \sigma^2$ and that for a twice differentiable function g we have

$$\left|g'(x+t)^2 - g'(x)^2\right| \leq 2\|g'g''\|\|t\|, \tag{2.3}$$

where $\|\cdot\|$ is the supremum norm. We note that $g'(W) = g'(X_1 + W^{(1)})$ with $W^{(1)} = W - n^{-1/2}X_1$, so that we can use the independence between X_1 and $W^{(1)}$ to deduce the upper bound

$$\text{Var}[g(W)] \leq \sigma^2 \mathbb{E}[(g'(W^{(1)}))^2] + \frac{C_U(g)}{\sqrt{n}},$$

where $C_U(g) = 2\|g'g''\|\sigma^2\mathbb{E}[|X_1|]$. *As we might expect from the CLT, we have obtained an upper bound with a form similar to that of* (1.1), *though with W replaced by $W^{(1)}$ and an additional term of order $O(n^{-1/2})$.*

Similar considerations apply for the lower bound. First recall that, if g is twice differentiable, then $g'(x+t) = g'(x) + t\mathbb{E}[g''(x+Ut)]$ with $U \sim \mathrm{Unif}[0,1]$. By independence, we can write

$$\mathbb{E}\left[\tau_1(X_1)g'(W)\right] = \sigma^2\mathbb{E}[g'(W^{(1)})] + n^{-1/2}\mathbb{E}[\tau_1(X_1)X_1g''(W^{(1)} + n^{-1/2}UX_1)],$$

so that the lower bound becomes

$$\mathrm{Var}[g(W)] \geq \sigma^2\mathbb{E}[g'(W^{(1)})]^2 + \frac{C_L(g,n)}{\sqrt{n}},$$

where

$$C_L(g,n) = 2\mathbb{E}[g'(W^{(1)})]\mathbb{E}[\tau_1(X_1)X_1g''(W^{(1)} + n^{-1/2}UX_1)] + \frac{n^{-1/2}}{\sigma^2}\mathbb{E}[\tau_1(X_1)X_1g''(W^{(1)} + n^{-1/2}UX_1)]^2.$$

Clearly $\lim_{n\to\infty} C_L(g,n)/\sqrt{n} = 0$ for all g, and so we have a lower bound similar in form to (1.2), *though again with $W^{(1)}$ replacing W in the main term, and an additional term which goes to zero as $n \to \infty$.*

Our second example yields variance bounds for any random variable with finite mean.

Example 2.2 (Smoothing). *Let Y be a real-valued random variable with $\mathbb{E}[Y] = \mu$. In order to allow us to derive variance bounds for Y using our approach, we smooth it by convolving it with Gaussian noise $Z \sim \mathcal{N}(0, \epsilon^2)$, independent of Y, for some small $\epsilon > 0$. Let φ_ϵ and Φ_ϵ be the density and distribution functions of Z, respectively, and define*

$$\tau_\epsilon(x) = \epsilon^2 + \frac{\mathbb{E}\left[(Y'-\mu)\bar{\Phi}_\epsilon(x-Y')\right]}{\mathbb{E}[\varphi_\epsilon(x-Y')]}, \tag{2.4}$$

where $\bar{\Phi}_\epsilon(y) = 1 - \Phi_\epsilon(y)$ and Y' is an independent copy of Y. Then $\tau_\epsilon(x)$ is a Stein kernel for $Y + Z$ (see Appendix A) and (2.2) *applies to all differentiable functions $g : \mathbb{R} \to \mathbb{R}$ such that $g(Y+Z)$ has finite second moment. Moreover, the following hold:*

(i). *If the mapping $x \mapsto (g(x) - \mathbb{E}[g(Y+Z)])^2$ is convex, then*

$$\mathrm{Var}[g(Y)] \leq \mathbb{E}\left[\tau_\epsilon(Y+Z)g'(Y+Z)^2\right].$$

(ii). *If the mapping $x \mapsto (g(x) - \mathbb{E}[g(Y)])^2$ is concave, then*

$$\mathrm{Var}\left[g(Y)\right] \geq \frac{\mathbb{E}\left[\tau_\epsilon(Y+Z)g'(Y+Z)\right]^2}{\epsilon^2 + \mathrm{Var}[Y]}.$$

We defer the proofs of these claims to Appendix A.

In our final set of examples in this section we illustrate how our techniques can be applied to the rich class of Pearson family distributions, particularly in the setting of Bayesian inference.

Example 2.3 (Pearson family and application to posterior distributions). *As is well known, the Pearson family has explicit Stein kernels given by Proposition B.1 recalled in the Appendix. Such a result is particularly useful in the following situation inherited from Bayesian statistics. In a Bayesian setting, the initial distribution of the parameter of interest is some prior distribution with density $\pi_0(\theta)$; upon observing data points $\mathbf{x} = (x_1, \ldots, x_n)$ sampled independently with sampling distribution $\pi(\theta, \mathbf{x})$ we update from the prior to the posterior density given by $\pi_2(\theta) = \kappa_2(\mathbf{x})\pi(\theta, \mathbf{x})\pi_0(\theta)$, where $\kappa_2(\mathbf{x})$ is a normalizing constant depending only on the data. We use the notation Θ_0 to indicate the distribution of the parameter under the prior, Θ_2 its distribution under the posterior, and X a random variable following the same common distribution of the observations. We also write Θ_1 for the parameter under the sampling distribution $\pi_1(\theta) = \kappa_1(\mathbf{x})\pi(\theta, \mathbf{x})$ with $\kappa_1(\mathbf{x})$ the corresponding normalizing constant, which corresponds to a posterior with flat (uninformative) prior. A popular choice of prior is that of a conjugate prior as the corresponding posterior follows the same distribution as the prior; the impact of the data is then visible in the parameters of the posterior distribution which are updated. Restricting our attention to Pearson distributed families, we can apply Proposition B.1 and read variance bounds directly from the updated parameters. For instance:*

- *Gaussian data, inference on mean, Gaussian prior: If $X \sim \mathcal{N}(\theta, \sigma^2)$ with $\theta \in \mathbb{R}$ and fixed $\sigma > 0$, and $\Theta_0 \sim \mathcal{N}(\mu, \delta^2)$ with $\mu \in \mathbb{R}, \delta > 0$, then $\Theta_2 \sim \mathcal{N}\left(\frac{\sigma^2 \mu + n \delta^2 \bar{x}}{n \delta^2 + \sigma^2}, \frac{\sigma^2 \delta^2}{n \delta^2 + \sigma^2}\right)$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. The Stein kernel for this Gaussian distribution is $\tau(\theta) = (\frac{n}{\sigma^2} + \frac{1}{\delta^2})^{-1}$. Consequently,*

$$\mathbb{E}\left[g'(\Theta_2)\right]^2 \leq \left(\frac{n}{\sigma^2} + \frac{1}{\delta^2}\right) \mathrm{Var}[g(\Theta_2)] \leq \mathbb{E}[g'(\Theta_2)^2]$$

  for all suitable g, all n and all values of the parameters.
- *Gaussian data, inference on variance, Inverse Gamma prior: If $X \sim \mathcal{N}(\mu, \theta)$ with $\theta > 0$ and fixed $\mu \in \mathbb{R}$, and $\Theta_0 \sim \mathcal{IG}(\alpha, \beta)$ has an Inverse Gamma distribution with density*

$$\theta \mapsto \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{-\alpha-1} \exp\left(-\frac{\beta}{\theta}\right), \, \alpha, \beta > 0,$$

  then $\Theta_2 \sim \mathcal{IG}\left(\frac{n}{2} + \alpha, \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 + \beta\right)$. The Stein kernel for this Inverse Gamma distribution is $\tau(\theta) = \frac{\theta^2}{\frac{n}{2} + \alpha - 1}$. Consequently, for all suitable g,

$$\frac{(\frac{n}{2} + \alpha - 2)}{(\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 + \beta)^2} \mathbb{E}[\Theta_2^2 g'(\Theta_2)]^2 \leq \mathrm{Var}[g(\Theta_2)] \leq \frac{1}{\frac{n}{2} + \alpha - 1} \mathbb{E}[\Theta_2^2 g'(\Theta_2)^2].$$

- *Binomial data, inference on proportion, Beta prior: If $X \sim Bin(n, \theta)$ with $\theta \in [0,1]$, and $\Theta_0 \sim Beta(\alpha, \beta)$ with density*

$$\theta \mapsto \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}, \, \alpha, \beta > 0,$$

  then $\Theta_2 \sim Beta\left(x + \alpha, n - x + \beta\right)$, where x denotes the observed number of successes. The Stein kernel for this Beta distribution is $\tau(\theta) = \frac{\theta(1-\theta)}{n+\alpha+\beta}$. Consequently, for all suitable g,

$$\frac{(n + \alpha + \beta + 1)}{(x + \alpha)(n - x + \beta)} \mathbb{E}[\Theta_2(1 - \Theta_2)g'(\Theta_2)]^2 \leq \mathrm{Var}[g(\Theta_2)] \leq \frac{\mathbb{E}[\Theta_2(1 - \Theta_2)g'(\Theta_2)^2]}{n + \alpha + \beta}.$$

In each example, we can see how the fixed parameters of the data as well as the fixed prior parameters interplay with the sample size and the data to yield bounds on the variance. Of course, one can state these bounds in a more general setting for each distribution for which we know the Stein kernel, but we choose to use the Bayesian setting since, to the best of our knowledge, these are the first variance bounds within a Bayesian context. Further examples are provided in Appendix B.

Remark 2.4. *A possible research direction springing from this last example, brought to our attention by the Associate Editor, is in the context of the famous Bernstein-von Mises theorem. More specifically, the data-dependent variance bounds may be used to provide rates of convergence of the posterior distribution towards the Gaussian distribution. This is an interesting research question, which deserves separate treatment.*

3. Variance bounds from zero-biased couplings

In this section, we suppose that the target random variable W has mean zero, finite variance σ^2, and can be coupled to some random variable $W^\star$ through

$$\mathbb{E}[W\phi(W)] = \sigma^2 \mathbb{E}[\phi'(W^\star)] \tag{3.1}$$

for all differentiable functions $\phi : \mathbb{R} \to \mathbb{R}$ such that $x\phi(x) \in L^1(W)$. Such $W^\star$ always exists, and its law is unique. It has the W-zero-biased distribution; see, e.g., [6, Section 2.3.3] and references therein for more details. $W^\star$ is a continuous random variable, regardless of whether W is discrete or continuous. Under (3.1), we immediately obtain

$$\sigma^2 \mathbb{E}\left[g'(W^\star)\right]^2 \leq \mathrm{Var}[g(W)] \leq \sigma^2 \mathbb{E}\left[g'(W^\star)^2\right] \tag{3.2}$$

by using (1.4) and (1.5) with $\gamma(x) = x$, $T_1 = \sigma^2$ and $T_2 = W^\star$ for all appropriate $g : \mathbb{R} \to \mathbb{R}$. Obviously it may be of interest to express (3.2) in terms of the original variable. By applying (2.3), we immediately obtain the following result.

Proposition 3.1. *Let W have mean zero and finite variance σ^2, and $W^\star$ have the W-zero biased distribution. Then*

$$\mathrm{Var}[g(W)] \leq \sigma^2 \mathbb{E}\left[g'(W)^2\right] + 2\sigma^2 \|g'g''\| \mathbb{E}|W^\star - W| \tag{3.3}$$

for all twice differentiable functions $g : \mathbb{R} \to \mathbb{R}$ for which $\mathrm{Var}[g(W)]$ exists.

It is classical that the Gaussian distribution is the unique fixed point of the zero-bias transform, in the sense that $W \sim \mathcal{N}(0, \sigma^2)$ if and only if $W = W^\star$. Hence $|W^\star - W|$ gives information on the distributional proximity between the law $\mathcal{L}(W)$ of W and $\mathcal{N}(0, \sigma^2)$. Also, it is classical that the Gaussian is characterized by the fact that $\sigma^2 = \sup_g \mathrm{Var}[g(W)]/\mathbb{E}[g'(W)^2]$; see, e.g., [4]. Inequality (3.3) captures these two essential features of the Gaussian distribution.

We conclude this section with two applications to illustrate the bound (3.3), the first to sums of independent random variables and the second to a combinatorial central limit theorem.

Example 3.2. *Let $X_1, X_2, \ldots, X_n$ be independent mean zero random variables with finite variances $\mathbb{E}[X_i^2] = \sigma_i^2, i = 1, \ldots, n$. Set $W = X_1 + \cdots + X_n$ and $\mathbb{E}[W^2] = \sigma^2 = \sum_{i=1}^n \sigma_i^2$. Let I be a random index independent of all else such that $P(I = i) = \sigma_i^2/\sigma^2$ and let $W_i = W - X_i$. Finally let $X_i^\star$ be the zero-bias transform of X_i. Then $W^\star - W = X_I - X_I^\star$ (see Lemma 2.1(v) of [11]) so that the bound (3.3) becomes*

$$\mathrm{Var}[g(W)] \leq \sigma^2 \mathbb{E}[g'(W)^2] + 2\|g'g''\| \sum_{i=1}^n \sigma_i^2 \mathbb{E}[|X_i - X_i^\star|].$$

If, furthermore, we suppose the summands to be independent copies of X such that $\sigma^2 = 1$ then

$$\mathrm{Var}[g(W)] \leq \mathbb{E}[g'(W)^2] + 2\|g'g''\| \mathbb{E}[|X - X^\star|].$$

To see how this plays out in practice, suppose that $X = (\xi - p)/\sqrt{npq}$ with ξ Bernoulli with success parameter p. Following [6, Corollary 4.1], we obtain $\mathbb{E}[|X - X^\star|] = (p^2 + q^2)/(2\sqrt{npq})$ and

$$\mathrm{Var}[g(W)] \leq \sigma^2 \mathbb{E}[g'(W)^2] + \|g'g''\| \frac{p^2 + q^2}{\sqrt{npq}}.$$

Many other examples can be explicitly worked out along these lines.

Example 3.3. *Let $(a_{i,j})_{i,j=1}^n$ be an array of real numbers and π a uniformly chosen permutation of $\{1, \ldots, n\}$. Let $W = \sum_{i=1}^n a_{i,\pi(i)}$. We further define*

$$a_{\bullet\bullet} = \frac{1}{n^2} \sum_{i,j=1}^n a_{i,j}, \quad a_{i\bullet} = \frac{1}{n} \sum_{j=1}^n a_{i,j}, \quad and \quad a_{\bullet j} = \frac{1}{n} \sum_{i=1}^n a_{i,j},$$

and note that $\mathbb{E}[W] = na_{\bullet\bullet}$ and

$$\mathrm{Var}[W] = \sigma^2 = \frac{1}{n-1} \sum_{i,j=1}^n (a_{i,j} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet})^2.$$

See, for example, [6, Section 4.4]. Letting $Z = \sigma^{-1}(W - na_{\bullet\bullet})$ and $C = \max_{1 \leq i,j \leq n} |a_{i,j} - a_{i\bullet} - a_{\bullet j} + a_{\bullet\bullet}|$, the proof of Theorem 6.1 of [6] shows that $\mathbb{E}|Z^\star - Z| \leq 8C\sigma^{-1}$, and so we have from (3.3) that

$$\mathrm{Var}[g(Z)] \leq \mathbb{E}\left[g'(Z)^2\right] + \frac{16C}{\sigma} \|g'g''\|,$$

for all appropriate g.

4. Variance bounds using stochastic ordering

We consider now some further applications in which we do not require explicit knowledge of the density of W in order to derive bounds on $\mathrm{Var}[g(W)]$ using our techniques. Unlike those examples in Section 3, the bounds we obtain here have the same form as in applications where we employ the exact expression for the underlying density, without any additional 'remainder' terms. We may obtain such bounds under natural assumptions on the random variable W, which we express in terms of stochastic orderings; the price we pay is in some restriction on the class of functions g for which the bounds apply.

We begin by recalling the definitions of the orderings which we will use. For any random variables X and Y, we will say that X is stochastically smaller than Y (denoted $X \leq_{st} Y$) if $\mathbb{P}(X > t) \leq \mathbb{P}(Y > t)$ for

all t. We will say that X is smaller than Y in the convex order (denoted $X \leq_{cx} Y$) if $\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)]$ for all convex functions ϕ for which the expectations exist. See [23] for background and many further details.

4.1. *Zero-biased couplings and the convex order.* Let W be a real-valued random variable with mean zero and variance σ^2. Recall the definition (3.1) of $W^\star$, the zero-biased version of W. We note that, from Lemma 2.1(ii) of [11], $W^\star$ is supported on the closed convex hull of the support of W and has density function given by

$$p_W^\star(w) = \frac{1}{\sigma^2} \mathbb{E}[WI(W > w)]. \tag{4.1}$$

Now, as in Section 3, we apply the upper bound (1.4) with the choices $\gamma(x) = x$, $T_1 = \sigma^2$ and $T_2 = W^\star$ to obtain $\mathrm{Var}[g(W)] \leq \sigma^2 \mathbb{E}[g'(W^\star)^2]$ for all appropriate g. If we assume that $W^\star \leq_{cx} W$ and that $g'(x)^2$ is convex, we can immediately bound $\mathbb{E}[g'(W^\star)^2]$ by $\mathbb{E}[g'(W)^2]$, and obtain the following result.

Theorem 4.1. *Let W have mean 0 and variance σ^2, and assume that $W^\star \leq_{cx} W$. Then*

$$\mathrm{Var}[g(W)] \leq \sigma^2 \mathbb{E}[g'(W)^2] \tag{4.2}$$

for all differentiable $g : \mathbb{R} \to \mathbb{R}$ such that $\mathrm{Var}[g(W)] < \infty$ and $x \mapsto g'(x)^2$ is convex.

The following example illustrates one setting where the ordering $W^\star \leq_{cx} W$ holds.

Example 4.2. *Let $W = X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent, mean-zero random variables, with X_i supported on the set $\{-a_i, b_i\}$ for $a_i, b_i > 0$, for each $i = 1, \ldots, n$. That is, $\mathbb{P}(X_i = -a_i) = p_i = 1 - \mathbb{P}(X_i = b_i)$ for $1 \leq i \leq n$, where $p_i = b_i/(a_i + b_i)$ so that $\mathbb{E}[X_i] = 0$. Let $\sigma_i^2 = \mathrm{Var}(X_i)$ and $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2$.*

A straightforward calculation using (4.1) shows that, for each $i = 1, \ldots, n$, $X_i^\star$ is uniformly distributed on the interval $[-a_i, b_i]$. Hence, Theorem 3.A.44 of [23] gives that $X_i^\star \leq_{cx} X_i$ for each i.

Let I be a random index, chosen independently of all else, with $\mathbb{P}(I = i) = \sigma_i^2/\sigma^2$, for $i = 1, \ldots, n$. Now, using Lemma 2.1(v) of [11], $W^\star$ is equal in distribution to $X_I^\star + \sum_{j \neq I} X_j$, which is smaller than W in the convex order by (3.A.46) of [23]. It then follows from Theorem 3.A.12(b) of [23] that $W^\star \leq_{cx} W$, and hence our upper bound (4.2) applies.

4.2. *Equilibrium couplings.* Throughout this section, let W be a non-negative random variable with mean λ^{-1}. Following, for example, [19], we say that a random variable W^e has the equilibrium distribution with respect to W if

$$\mathbb{E}[\phi(W)] - \phi(0) = \lambda^{-1} \mathbb{E}[\phi'(W^e)], \tag{4.3}$$

for all differentiable functions $\phi : \mathbb{R}^+ \to \mathbb{R}$.

Remark 4.3. *This definition is motivated by the fact that W is Exponential if and only if W and W^e are equal in distribution. Applying the definition to the function $\phi_x(w) = (w-x)\mathbb{I}(w \geq x)$ and integrating by parts we obtain that $\mathbb{P}(W^e > x) = \lambda \int_x^\infty \mathbb{P}(W > y)\, dy$ for all $x \geq 0$.*

In this section we consider random variables that are new better than used in expectation (NBUE) and new worse than used in expectation (NWUE). Recall that W is NBUE if

$$\frac{\int_x^\infty \mathbb{P}(W > s)\, ds}{\int_0^\infty \mathbb{P}(W > s)\, ds} = \lambda \int_x^\infty \mathbb{P}(W > s)\, ds \leq \mathbb{P}(W > x) \tag{4.4}$$

for all $x \geq 0$, and that W is NWUE if (4.4) holds with the inequality reversed. These properties are well-known in reliability theory and are closely connected with many other classes of distributions frequently used in reliability analysis; see, for example, [23]. We note, for example, that distributions with increasing hazard rate are NBUE and those with decreasing hazard rate are NWUE, and so the results of this section may be applied to distributions with monotone hazard rate. Note also that the definition (4.4) of the NBUE property can be equivalently expressed as $\mathbb{E}[W - t | W > t] \leq \mathbb{E}[W]$ for all t. Defining the conditional value at risk (CVaR) of W as $\mathbb{E}[W | W > Q_W(u)]$, where $Q_W(u) = \inf\{x : \mathbb{P}(W > x) \leq u\}$ is the quantile function of W, the random variable W is therefore NBUE if its CVaR is bounded above by the CVaR of an exponential distribution with the same mean. Recent applications of CVaR to

stochastic inequalities include Gaussian approximation results for CVaR of sums of random variables [20] and exponential inequalities for suprema of heavy-tailed empirical processes [16].

From the definition (4.4) and the remark above it, it is clear that W is NBUE if and only if $W^e \leq_{st} W$, and that W is NWUE if and only if $W \leq_{st} W^e$. For a random variable W which is either NBUE or NWUE, we employ this stochastic ordering in a similar way to the convex ordering we used in Section 4.1 above.

We will seek conditions on the random variable W and the differentiable function ϕ such that

$$\mathbb{E}[(\lambda W - 1)\phi(W)] \leq \mathbb{E}[W\phi'(W)]. \tag{4.5}$$

This inequality is analogous to the equality (1.3) we used in deriving the upper bound (1.4), with the choices $\gamma(x) = \lambda x - 1$ and $T_1 = T_2 = W$. Following the proof of (1.4), we see that such an inequality is sufficient to establish an upper variance bound; the equality we used in (1.3) is not necessary. Now, the definition (4.3) of W^e gives that

$$\mathbb{E}[W\phi(W)] = \lambda^{-1}\mathbb{E}[\phi(W^e) + W^e\phi'(W^e)],$$

and hence

$$\mathbb{E}[(\lambda W - 1)\phi(W)] + \mathbb{E}[\phi(W)] = \mathbb{E}[W^e\phi'(W^e)] + \mathbb{E}[\phi(W^e)].$$

Thus, (4.5) holds if and only if

$$\mathbb{E}[\phi(W^e) + W^e\phi'(W^e)] \leq \mathbb{E}[\phi(W) + W\phi'(W)].$$

Therefore, inequality (4.5) holds if W is NBUE and $\phi(x) + x\phi'(x)$ is increasing in x. Alternatively, (4.5) also holds if W is NWUE and $\phi(x) + x\phi'(x)$ is decreasing in x.

Continuing to follow the proof of the upper bound (1.4), we now apply (4.5) with the choice

$$\phi(x) = \int_0^{\gamma(x)} \left(\frac{g'(\gamma^{-1}(u))}{\gamma'(\gamma^{-1}(u))}\right)^2 \mathrm{d}u = \frac{1}{\lambda^2}\int_0^{\lambda x - 1} g'(\lambda^{-1}(u+1))^2\mathrm{d}u,$$

so that we may conclude our argument with an upper bound on $\mathrm{Var}[g(W)]$ for some differentiable function $g : \mathbb{R}^+ \to \mathbb{R}$. If g is in fact twice differentiable, we can check, for example, that this choice of ϕ gives $\phi(x) + x\phi'(x)$ increasing if $g'(x)[xg''(x) + g'(x)] \geq 0$ for all x, and decreasing if this inequality is reversed. We give a precise statement of the resulting upper variance bound in this setting in Theorem 4.4 below.

Similarly, we may ask when (4.5) holds with the inequality sign reversed; we can apply this reversed inequality in place of the equality (1.3) in a proof of a lower variance bound analogous to (1.5). By the same argument as above, (4.5) holds with the inequality reversed if either (i) W is NBUE and $\phi(x) + x\phi'(x)$ is decreasing in x, or (ii) W is NWUE and $\phi(x) + x\phi'(x)$ is increasing in x. Following the proof of (1.5) from (1.3), we can obtain a lower bound on $\mathrm{Var}[g(W)]$ for appropriate functions g by applying (4.5) with the inequality sign reversed and with the choice $\phi = g$.

We have thus proved the following, where we have restricted our attention to twice differentiable functions g for clarity in the stated results.

Theorem 4.4. *Let W be a non-negative random variable with mean $\mathbb{E}[W] = \lambda^{-1}$. Let $g : \mathbb{R}^+ \to \mathbb{R}$ be a twice differentiable function such that $\mathrm{Var}[g(W)]$ exists.*

(a) *Assume that either*
   (i) *W is NBUE and $g'(x)[xg''(x) + g'(x)] \geq 0$ for all $x \geq 0$; or*
   (ii) *W is NWUE and $g'(x)[xg''(x) + g'(x)] \leq 0$ for all $x \geq 0$.*
   Then
$$\mathrm{Var}[g(W)] \leq \frac{1}{\lambda}\mathbb{E}[Wg'(W)^2].$$

(b) *Assume that either*
   (i) *W is NBUE and $xg''(x) + 2g'(x) \leq 0$ for all $x \geq 0$; or*
   (ii) *W is NWUE and $xg''(x) + 2g'(x) \geq 0$ for all $x \geq 0$.*
   Then
$$\mathrm{Var}[g(W)] \geq \frac{(\mathbb{E}[Wg'(W)])^2}{\lambda^2\mathrm{Var}[W]}.$$

We conclude with an example illustrating the application of Theorem 4.4.

Example 4.5. *Consider the random sum $W = \sum_{i=1}^{N} X_i$, where $X, X_1, X_2, \ldots$ are independent and identically distributed, continuous, real-valued random variables and N is a random variable supported on the non-negative integers. Conditions are known under which W is NWUE. For example, [2] shows that if N is Geometric, then W is NWUE, regardless of the distribution of X. More generally, Corollary 2.1 of [25] establishes that if N satisfies*

$$\sum_{k=0}^{\infty} \mathbb{P}(N > n + k + 1) \geq \mathbb{P}(N > n) \sum_{k=0}^{\infty} \mathbb{P}(N > k), \tag{4.6}$$

for all $n = 0, 1, \ldots$, then W is NWUE. This includes, for example, the case where N is mixed Poisson with a mixing distribution that is itself NWUE; see Corollary 3.1 of [25]. Thus, under the condition (4.6), the bounds of the NWUE cases of Theorem 4.4 apply, with $\lambda^{-1} = \mathbb{E}[N]\mathbb{E}[X]$ and $\mathrm{Var}[W] = (\mathbb{E}[X])^2 \mathrm{Var}[N] + \mathbb{E}[N]\mathrm{Var}[X]$.

References

[1] M. Bonnefont, A. Joulin, and Y. Ma. A note on spectral gap and weighted Poincaré inequalities for some one-dimensional diffusions. *ESAIM: Probability and Statistics*, 20, 18–29, 2016.

[2] M. Brown. Error bounds for exponential approximations of geometric convolutions. *The Annals of Probability*, 18(3):1388–1402, 1990.

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

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

[5] L. H. Y. Chen. An inequality for the multivariate normal distribution. *Journal of Multivariate Analyis*, 12:306–315, 1982.

[6] 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.

[7] L. H. Y. Chen, and A. Röllin. Stein couplings for normal approximation. Preprint: https://arxiv.org/abs/1003.6039, 2010.

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

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

[10] M. Ernst, G. Reinert, and Y. Swan. On infinite covariance expansions. Preprint. Available at arXiv:1906.08376, 2019.

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

[12] L. Goldstein and G. Reinert. Distributional transformations, orthogonal polynomials, and Stein characterizations. *Journal of Theoretical Probability* 18(1):237–260, 2005.

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

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

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

[16] A. Marchina. An exponential inequality for suprema of empirical processes with heavy tails on the left *Comptes Rendus de l'Académie des Sciences de Paris, Ser. I*, 357(6): 537–544, 2019.

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

[18] I. Nourdin, G. Peccati, and Y. Swan. Integration by parts and representation of information functionals. *IEEE International Symposium on Information Theory (ISIT)*, pages 2217–2221, 2014.

[19] E. Peköz and A. Röllin. New rates for exponential approximation and the theorems of Rényi and Yaglom. *The Annals of Probability*, 39(2):587–608, 2011.

[20] E. Rio. About the conditional value at risk of partial sums *Comptes Rendus de l'Académie des Sciences de Paris, Ser. I*, 355(11): 1190–1195, 2017.

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

[22] A. Saumard and J.A. Wellner. Efron's monotonicity property for measures on $\mathbb{R}^2$. *Journal of Multivariate Analysis* 166: 212-224, 2018.

[23] M. Shaked and J. G. Shanthikumar. *Stochastic Orders.* Springer New York, 2007.

[24] C. Stein. *Approximate Computation of Expectations.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986.

[25] G. E. Willmot, S. Drekic, and J. Cai. Equilibrium compound distributions and stop-loss moments. *Scandinavian Actuarial Journal*, 2005(1):6–24, 2005.

Appendix A. Example 2.2: Proofs of claims

We begin by showing that $\tau_\epsilon(x)$, as defined in (2.4), is the Stein kernel of $Y + Z$. To see this, note that $\mathbb{P}(Y + Z \le t) = \mathbb{E}[\Phi_\epsilon(t - Y)]$, so that $Y + Z$ has density $p_\epsilon(t) = \mathbb{E}[\varphi_\epsilon(t - Y)]$. Hence, since $Y + Z$ has expectation μ, its Stein kernel is given by

$$\frac{1}{p_\epsilon(x)} \int_x^\infty (y - \mu) p_\epsilon(y) \, dy = \frac{1}{p_\epsilon(x)} \int_x^\infty \int_{-\infty}^\infty (y - \mu) \varphi_\epsilon(y - t) \, dF(t) \, dy \,,$$

where F is the distribution function of Y; see [4]. Applying Fubini's theorem, this is equal to

$$\frac{1}{p_\epsilon(x)} \int_{-\infty}^\infty \int_{x-t}^\infty (s + t - \mu) \varphi_\epsilon(s) \, ds \, dF(t) = \frac{1}{p_\epsilon(x)} \mathbb{E}\left[\epsilon^2 \varphi_\epsilon(x - Y) + (Y - \mu) \bar{\Phi}_\epsilon(x - Y)\right] \,,$$

since $\int_y^\infty s \varphi_\epsilon(s) \, ds = \epsilon^2 \varphi_\epsilon(y)$. This Stein kernel is easily seen to be equal to $\tau_\epsilon(x)$ given in (2.4).

Now, to prove claim (i), we firstly note that $Y \le_{cx} Y + Z$ (see Theorem 3.A.34 of [23]), so that $\mathbb{E}[\phi(Y)] \le \mathbb{E}[\phi(Y + Z)]$ for any convex function ϕ. Noting that the function $f(\alpha) = \mathbb{E}[(g(Y) - \alpha)^2]$ is minimized at $\alpha = \mathbb{E}[g(Y)]$, we have

$$\text{Var}[g(Y)] = \mathbb{E}\left[(g(Y) - \mathbb{E}[g(Y)])^2\right] \le \mathbb{E}\left[(g(Y) - \mathbb{E}[g(Y + Z)])^2\right] \le \text{Var}[g(Y + Z)] \,,$$

where the final inequality follows from the assumption in (i) that the mapping $x \mapsto (g(x) - \mathbb{E}[g(Y + Z)])^2$ is convex. Applying the upper bound from (2.2) completes the proof of (i).

We use a similar argument for (ii). We have that

$$\text{Var}[g(Y + Z)] \le \mathbb{E}[(g(Y + Z) - \mathbb{E}[g(Y)])^2] \le \mathbb{E}[(g(Y) - \mathbb{E}[g(Y)])^2] \,,$$

where the final inequality uses the convex ordering between Y and $Y + Z$ (from which $\mathbb{E}[\phi(Y + Z)] \le \mathbb{E}[\phi(Y)]$ for any concave function ϕ) and the assumption that the mapping $x \mapsto (g(x) - \mathbb{E}[g(Y)])^2$ is concave. We now apply the lower bound from (2.2) to complete the proof of (ii).

Appendix B. Example 2.3: Stein kernel and further applications

We start by recalling a result taken from [24, Equation (40), p. 65], which was used in Example 2.3.

Proposition B.1 (Pearson distribution). *A random variable with mean μ and variance σ^2 is of Pearson type if and only if there exist $\delta_1, \delta_2, \delta_3 \in \mathbb{R}$, not all equal to 0, such that*

$$\frac{p'(x)}{p(x)} = -\frac{(2\delta_1 + 1)(x - \mu) + \delta_2}{\delta_1(x - \mu)^2 + \delta_2(x - \mu) + \delta_3}.$$

In this case, its Stein kernel is $\tau(x) = \delta_1(x - \mu)^2 + \delta_2(x - \mu) + \delta_3$.

To complement Example 2.3 and illustrate the scope of its application, we use the remainder of this appendix to present further examples along similar lines.

Example B.2 (Negative binomial data, inference on proportion, Beta prior)**.** *If $X \sim NB(r, \theta)$ has a negative binomial distribution with $\theta \in [0, 1]$ and fixed $r \in \mathbb{N}$, and $\Theta_0 \sim Beta(\alpha, \beta)$ with $\alpha, \beta > 0$, then $\Theta_2 \sim Beta\left(\sum_{i=1}^{n} x_i + \alpha, nr + \beta\right)$. The Stein kernel for this Beta distribution is $\tau(\theta) = \frac{\theta(1-\theta)}{\sum_{i=1}^{n} x_i + nr + \alpha + \beta}$. Consequently,*

$$\frac{(\sum_{i=1}^{n} x_i + nr + \alpha + \beta + 1)}{(\sum_{i=1}^{n} x_i + \alpha)(nr + \beta)} \mathbb{E}[\Theta_2 (1 - \Theta_2) g'(\Theta_2)]^2 \leq \mathrm{Var}[g(\Theta_2)] \leq \frac{\mathbb{E}[\Theta_2 (1 - \Theta_2) g'(\Theta_2)^2]}{\sum_{i=1}^{n} x_i + nr + \alpha + \beta}.$$

Example B.3 (Weibull data, inference on scale, Inverse Gamma prior)**.** *If $X \sim Wei(k, \theta)$ has a Weibull distribution with $\theta > 0$ and fixed $k > 0$ (note that here we consider the Weibull density $x \mapsto \frac{kx^{k-1}}{\theta} \exp(-x^k/\theta), x > 0$), and $\Theta_0 \sim IG(\alpha, \beta)$ with $\alpha, \beta > 0$, then $\Theta_2 \sim IG\left(n + \alpha, \sum_{i=1}^{n} x_i^k + \beta\right)$. The Stein kernel for this Inverse Gamma distribution is $\tau(\theta) = \frac{\theta^2}{n + \alpha - 1}$. Consequently,*

$$\frac{n + \alpha - 2}{(\sum_{i=1}^{n} x_i^k + \beta)^2} \mathbb{E}[\Theta_2^2 g'(\Theta_2)]^2 \leq \mathrm{Var}[g(\Theta_2)] \leq \frac{\mathbb{E}[\Theta_2^2 g'(\Theta_2)^2]}{n + \alpha - 1}.$$

Example B.4 (Gamma data, inference on scale, Gamma prior)**.** *If $X \sim Gam(k, \theta)$ has a Gamma distribution with $\theta, k > 0$, and $\Theta_0 \sim Gam(\alpha, \beta)$ with $\alpha, \beta > 0$, then $\Theta_2 \sim Gam\left(nk + \alpha, \sum_{i=1}^{n} x_i + \beta\right)$. The Stein kernel for this Gamma distribution is $\tau(\theta) = \frac{\theta}{\sum_{i=1}^{n} x_i + \beta}$. Consequently,*

$$\frac{\mathbb{E}[\Theta_2 g'(\Theta_2)]^2}{nk + \alpha} \leq \mathrm{Var}[g(\Theta_2)] \leq \frac{1}{\sum_{i=1}^{n} x_i + \beta} \mathbb{E}[\Theta_2 g'(\Theta_2)^2].$$

Example B.5 (Laplace data, inference on scale, inverse gamma prior)**.** *If $X \sim Lap(\mu, \theta)$ has a Laplace distribution with $\theta > 0$ and fixed $\mu \in \mathbb{R}$, and $\Theta_0 \sim IG(\alpha, \beta)$ with $\alpha, \beta > 0$, then we have that $\Theta_2 \sim IG\left(n + \alpha, \sum_{i=1}^{n} |x_i - \mu| + \beta\right)$. The Stein kernel can readily be deduced from previous examples, and we get*

$$\frac{n + \alpha - 2}{(\sum_{i=1}^{n} |x_i - \mu| + \beta)^2} \mathbb{E}\left[\Theta_2^2 g'(\Theta_2)\right]^2 \leq \mathrm{Var}[g(\Theta_2)] \leq \frac{1}{n + \alpha - 1} \mathbb{E}\left[\Theta_2^2 g'(\Theta_2)^2\right].$$

Example B.6 (Poisson data, inference on mean=scale, Gamma prior)**.** *If $X \sim Poi(\theta)$ has a Poisson distribution with $\theta > 0$, and $\Theta_0 \sim Gam(\alpha, \beta)$ with $\alpha, \beta > 0$, then $\Theta_2 \sim Gam\left(\sum_{i=1}^{n} x_i + \alpha, n + \beta\right)$. The Stein kernel can readily be deduced from previous examples, and we get*

$$\frac{\mathbb{E}[\Theta_2 g'(\Theta_2)]^2}{\sum_{i=1}^{n} x_i + \alpha} \leq \mathrm{Var}[g(\Theta_2)] \leq \frac{1}{n + \beta} \mathbb{E}\left[\Theta_2 g'(\Theta_2)^2\right].$$

Example B.7 (Uniform data, inference on interval length, Pareto prior)**.** *If $X \sim U(0, \theta)$ has a Uniform distribution with $\theta > 0$, and $\Theta_0 \sim Par(\alpha, \beta)$ has a Pareto distribution with $\alpha, \beta > 0$ (as a reminder, the density of such a Pareto distribution is $\theta \mapsto \frac{\alpha \beta^\alpha}{\theta^{\alpha+1}} \mathbb{I}[\beta \leq \theta]$ where $\mathbb{I}[A]$ is the indicator function of the event A), then $\Theta_2 \sim Par\left(n + \alpha, \max(m(x), \beta)\right)$ with $m(x) = \max(x_1, \ldots, x_n)$. The Stein kernel for this Pareto distribution is $\tau(\theta) = \frac{\max(m(x), \beta) - \theta}{n + \alpha - 1} \theta$. Consequently, we get*

$$\frac{(n + \alpha - 2)}{(n + \alpha)(\max(m(x), \beta))^2} \mathbb{E}\left[(\max(m(x), \beta) - \Theta_2) \Theta_2 g'(\Theta_2)\right]^2$$

$$\leq \mathrm{Var}[g(\Theta_2)] \leq \frac{1}{n + \alpha - 1} \mathbb{E}\left[(\max(m(x), \beta) - \Theta_2) \Theta_2 g'(\Theta_2)^2\right].$$