An algebra of Stein operators

Robert E. Gaunt[*], Guillaume Mijoule [†] and Yvik Swan[‡]

The University of Manchester, INRIA and Université de Liège

Abstract

We build upon recent advances on the distributional aspect of Stein's method to propose a novel and flexible technique for computing Stein operators for random variables that can be written as products of independent random variables. We show that our results are valid for a wide class of distributions including normal, beta, variance-gamma, generalized gamma and many more. Our operators are kth degree differential operators with polynomial coefficients; they are straightforward to obtain even when the target density bears no explicit handle. As an application, we derive a new formula for the density of the product of k independent symmetric variance-gamma distributed random variables.

AMS classification: Primary 60E15; Secondary 62E15

Key words: Stein's method, Stein operators, product distributions, variance-gamma distribution.

1 Introduction

In 1972, Charles Stein (1920–2016) [41] introduced a powerful method for estimating the error in normal approximations. The method was adapted to the Poisson distribution by Louis Chen in [8], and has since been extended to a very broad family of probability distributions. The general procedure for a given target distribution p is as follows. In the first step, one obtains a suitable operator A acting on a class of test functions $\mathcal{F}$ such that $\mathbb{E}[Af(X)] = 0$ for all $f \in \mathcal{F}$; the operator A is called a *Stein operator* for p. For continuous distributions, A is a typically a differential operator; for the standard normal distribution, the classical operator is $Af(x) = f'(x) - xf(x)$. One then considers the so-called *Stein equation*

$$Af_h(x) = h(x) - \mathbb{E}h(X), \tag{1}$$

where h is a real-valued test function. If A is well chosen then, for a given h, the Stein equation (1) can be solved for f_h. The second step of the method consists of obtaining this solution and then bounding appropriate lower order derivatives. Evaluating both sides of (1) at a random

[*]School of Mathematics, The University of Manchester, Manchester M13 9PL, UK robert.gaunt@manchester.ac.uk

[†]MOKAPLAN, INRIA de Paris, 2 rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France guillaume.mijoule@inria.fr.

[‡]Université de Liège, Sart-Tilman, Allée de la découverte 12, B-8000 Liège, Belgium yswan@uliege.be.

variable of interest W and taking the supremum over all h in some class of functions $\mathcal{H}$ leads to the estimate

$$d_{\mathcal{H}}(\mathcal{L}(W), \mathcal{L}(X)) := \sup_{h\in\mathcal{H}} |\mathbb{E}h(W) - \mathbb{E}h(X)| \leq \sup_{f_h} |\mathbb{E}[Af_h(W)]|, \tag{2}$$

where the final supremum is taken over all f_h that solve (1). The third and final step of the method involves developing appropriate strategies for bounding the expectation on the right hand side of (2). This is of interest because many important probability metrics (such as the Kolmogorov and Wasserstein metrics) are of the form $d_{\mathcal{H}}(\mathcal{L}(W), \mathcal{L}(X))$. Moreover, in many settings bounding the expectation $\mathbb{E}[Af_h(W)]$ is relatively tractable, and as a result Stein's method has found application in disciplines as diverse as random graph theory [5], number theory [22], statistical mechanics [13] and quantum mechanics [29]. We refer to the survey paper [37] as well as to the monographs [30, 9] for a deeper look into some of the fruits of Charles Stein's seminal insights, particularly in the case where the target is the normal distribution.

The linchpin of the method is the operator A whose properties are crucial to the success of the whole enterprise. In the sequel, we concentrate exclusively on *differential* Stein operators (some operators in the literature are integral or even fractional, see e.g. [43, 3]) and adopt the following lax definition:

Definition 1.1. *A linear differential operator A acting on a class $\mathcal{F}$ of functions is a* Stein operator *for X if (i) $Af \in L^1(X)$ and (ii) $\mathbb{E}[Af(X)] = 0$ for all $f \in \mathcal{F}$.*

There are infinitely many Stein operators for any given target distribution. For instance, if the distribution is known (even if only up to a normalizing constant) then the "canonical" theory from [26] applies, leading to entire families of operators. This approach provides natural first order polynomial operators e.g. for target distributions which belong to the Pearson family [38] or which satisfy a diffusive assumption [11, 25]. In some cases, one may rather apply a duality argument. For instance the p.d.f. $\gamma(x) = (2\pi)^{-1/2}e^{-x^2/2}$ of the standard normal distribution satisfies the first order ODE $\gamma'(x) + x\gamma(x) = 0$ leading, by integration by parts, to the already mentioned operator $Af(x) = f'(x) - xf(x)$. This is particularly useful for densities defined implicitely via ODEs. Such are by no means the only methods for deriving differential Stein operators and, for any given X, one can easily determine an entire ecosystem of Stein operators, leading to the natural question of *which operator to choose.* One natural way to sieve through the available options is to further impose that the chosen operator be characterizing for X, i.e. that if some Y enjoys the property that $\mathbb{E}[f(Y)] = 0$ for all $f \in \mathcal{F}$, then $Y{=}X$ (equality in law). Such requirements often do not suffice and will not be imposed here; our focus will rather be on another crucial quality of a "good" Stein operator: tractability. More precisely, we will focus solely on Stein operators which satisfy the next definition.

Definition 1.2. *We call a Stein operator* polynomial *if it can be written as a finite sum $A = \sum_{i,j} a_{ij}M^iD^j$ for real coefficients $a_{ij} \in \mathbb{R}$, with $M(f) = (x \mapsto xf(x))$ and $D(f) = (x \mapsto f'(x))$.*

Except in the most basic cases, determining polynomial Stein operators is not an easy task. Interestingly, many densities do not admit a first order polynomial Stein operator and it is necessary to consider higher order operators: [15] obtains a second order operator for the entire family of variance-gamma distributions (see also [14] and [17]), [36] obtain a second order Stein operator for the Laplace distribution, and [34] obtain a second order operator for the PRR distribution, which has a density that can be expressed in terms of the Kummer U function. If the p.d.f. of X is defined in terms of special functions (Kummer U, Meijer G, Bessel, etc.) which are themselves defined as solutions to explicit dth order differential equations then the duality approach shall yield a tractable differential operator with explicit coefficients.

In many cases, the target distribution is not even defined analytically in terms of its distribution but rather probabilistically, as a statistic (sum, product, quotient) of independent contributions. Explicit knowledge of the density of such random variables is then generally unavailable and, in order to obtain polynomial Stein operators for such objects, new approaches must be devised. In [2, 1], a Fourier-based approach is developed for identifying appropriate operators for arbitrary combinations of independent chi-square distributed random variables. In [19, 16], an iterative conditioning argument is provided for obtaining operators for (mixed) products of independent random beta, gamma and mean-zero normal random variables. In this context we are naturally lead to the following research problem:

> *Given independent random variables $X_1, \ldots, X_d$ with polynomial operators $A_1, \ldots, A_d$, respectively, can one deduce a tractable polynomial Stein operator for statistics of the form $X = F(X_1, \ldots, X_d)$?*

In this paper, we provide an answer for functionals of the form $F(x_1, \ldots, x_d) = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ with $\alpha_i \in \mathbb{R}$ and X_i's with polynomial Stein operator satisfying a specific commutativity assumption (Assumption 3 below).

This paper is mostly devoted to the problem of obtaining Stein operators, which allows for a focused treatment of their theory. We acknowledge, though, that further work is required before the Stein operators obtained in this paper can be used to prove approximation theorems via Stein's method. However, the theory of Stein operators to which this paper contributes is of importance in its own right. Indeed, there are now a wide variety of techniques which allow one to obtain useful bounds on solutions to the resulting Stein equations (see, for example, [24, 12, 3]) and which can be adapted to the operators that we derive. Also, and this has now been demonstrated in several papers such as [31, 32, 2, 4, 1], Stein operators can be used for comparison of probability distributions directly without the need of solving Stein equations; such an area is also the object of much interest. Finally, we stress that Stein operators are also of use in applications beyond proving approximation theorems; for example, in obtaining distributional properties [16, 19] and other surprising applications include the derivation of formulas for definite integrals of special functions [18]. Indeed, in Section 4, we propose a novel general technique for obtaining formulas of densities of distributions that may be intractable through other existing methods.

The outline of the paper is as follows. In Section 2 we identify the key elements allowing to construct a form of "operator algebra" which provides – by elementary calculations – polynomial operators for X's which can be written as products (see Section 2.4) and powers (see Section 2.5) of independent contributions. We apply the theory in Section 3 to recover several operators from contemporary literature on Stein's method and also to provide many new ones. Finally, in Section 4 we consider an application of operators obtained by our method to finding densities of product distributions.

2 An algebra of Stein operators

2.1 About the class of functions on which the operators are defined

All random variables we consider in the paper satisfy the following assumption:

Assumption 1: X admits a smooth density p with respect to the Lebesgue measure on $\mathbb{R}$; this density is defined and non-vanishing on some (possibly unbounded) interval $J \subseteq \mathbb{R}$.

By definition, a Stein operator A for a random variable X acts on a collection $\mathcal{F}$ of functions for which the expectations vanish. Although determining the largest possible set $\mathcal{F}$ may be an interesting quest, it will not be part of ours because this can only be done on a case-by-case basis and the focus of our paper is the construction of an algebra allowing to generate tractable operators. In order to ensure that the operators we obtain do not act on trivial classes of functions (e.g. $\mathcal{F} = \{\emptyset\}$), we shall simply impose the following assumption:

Assumption 2: X admits an operator A acting on $\mathcal{F}$ which contains the set of smooth functions with compact support $\mathcal{C}_0^\infty(\mathbb{R})$.

Assumption 2 is not too restrictive in our context (see Remark 2.2 below), although we will need to reinforce it slightly in Section 2.5. A collateral benefit of restricting to random variables satisfying Assumptions 1 and 2 is that we now may consider samples $X_1, \ldots, X_n$ of random variables with respective operators $A_1, \ldots, A_n$ acting on their respective classes $\mathcal{F}_1, \ldots, \mathcal{F}_n$ and we are ensured that $\bigcap_{i=1}^n \mathcal{F}_i \supseteq C_0^\infty(\mathbb{R})$. Consequentially, it is guaranteed that any statements on the joint behaviour of any of the A_i will also hold on non-trivial classes of functions.

Remark 2.1. *In some cases (for instance when X has exponential moments), one can easily extend $\mathcal{F}$ to smooth functions f such that $f^{(k)}$ has at most polynomial growth for all $k \geq 0$ (and in particular, to polynomials).*

Remark 2.2. *Under Assumption 1, the "canonical" Stein operator in [26] is $A_c : f \mapsto (fp)'/p = f' + p'/pf$ and we have $\mathbb{E}[A_c f(X)] = 0$ at least for those functions f such that $f(x)p(x) \to 0$ on the border of J. If $J = (-\infty, +\infty)$, it is clear that any $f \in \mathcal{C}_0^\infty(\mathbb{R})$ satisfies $f(x)p(x) \to 0$ on the border of J, hence Assumption 2 is automatically satisfied. On the other hand, if J has a finite border, say $J = (a, +\infty)$ with $a \in \mathbb{R}$, then $f \in \mathcal{C}_0^\infty(\mathbb{R})$ does not imply necessarily that fp vanishes on the border (see [7] for the case of exponential approximation). An easy workaround in this case is to apply the operator to $(x-a)f$ instead of f, which leads to the new operator $A = A_c(M - aI)$. This operator satisfies $\mathbb{E}[Af(X)] = 0$ for all $f \in \mathcal{C}_0^\infty(\mathbb{R})$. Of course, this has to be done on both borders of the support if they are both are finite.*

2.2 The building blocks

Let us record some notation regarding the different operators that will be used throughout the paper. We let $\mathcal{F} \supseteq \mathcal{C}_0^\infty(\mathbb{R})$; M is the multiplication operator: $M(f) = (x \mapsto xf(x))$; D the differentiation operator $D(f) = f'$; I the identity of $\mathcal{F}$; for $a \in \mathbb{R} \setminus \{0\}$, $\tau_a(f) = (x \mapsto f(ax))$; and $\forall r \in \mathbb{R}, T_r = MD + rI$.

Remark 2.3. *We will also need to consider the limit of operator T_r as $r \to \infty$. Although such a limit is badly defined, we note how $\lim_{r\to\infty} r^{-1}T_r = I$ (pointwisely for any $f \in \mathcal{F}$). Abusing notations, we will write $T_\infty = I$.*

Our starting point is the following extension of one of the main results of [19]:

Proposition 2.4. *Let $\mathcal{F} \supseteq \mathcal{C}_0^\infty(\mathbb{R})$. Assume X, Y are random variables with respective Stein operators*

$$A_X = L_X - M^p K_X, \tag{3}$$
$$A_Y = L_Y - M^p K_Y, \tag{4}$$

where $p \in \mathbb{N}$ and where the operators L_X, K_X, L_Y, K_Y commute with each other and with every τ_a, $a \in \mathbb{R}$. Then, if X and Y are independent,

$$L_X L_Y - M^p K_X K_Y \tag{5}$$

is a Stein operator for XY.

Proof. Let $f \in \mathcal{F}$. Using a conditioning argument and the commutative property between the different operators, we have that

$$\begin{aligned}
\mathbb{E}[L_X L_Y f(XY)] &= \mathbb{E}[\mathbb{E}[\tau_Y L_X L_Y f(X)\,|Y]] \\
&= \mathbb{E}[\mathbb{E}[L_X \tau_Y L_Y f(X)\,|Y]] \\
&= \mathbb{E}[\mathbb{E}[M^p K_X \tau_Y L_Y f(X)\,|Y]] \\
&= \mathbb{E}[\mathbb{E}[X^p \tau_Y K_X L_Y f(X)\,|Y]] \\
&= \mathbb{E}[X^p K_X L_Y f(XY)] \\
&= \mathbb{E}[X^p \mathbb{E}[\tau_X K_X L_Y f(Y)\,|\,X]] \\
&= \mathbb{E}[X^p \mathbb{E}[L_Y \tau_X K_X f(Y)\,|\,X]] \\
&= \mathbb{E}[X^p \mathbb{E}[M^p K_Y \tau_X K_X f(Y)\,|\,X]] \\
&= \mathbb{E}[X^p Y^p K_Y K_X f(XY)\,],
\end{aligned}$$

which achieves the proof. □

The assumption that the operators commute with scaling τ_a is crucial for the proof of Proposition 2.4; it will also reveal itself to be the linchpin of our "operator algebra". Restricting to first order differential operators we deduce that the fundamental operators L_X, K_X, L_Y, K_Y need to be of the form $T_r = MD + rI$, as these are the only first order polynomial Stein operators which commute with the multiplication operator (at least for a non trivial class $\mathcal{F}$). A fundamental subalgebra of linear operators, which will play a prominent role in this work, is the algebra $\mathcal{T}$ composed of all linear combinations and compositions of T_r's for $r \in \mathbb{R} \cup \{\infty\}$. Note also that $\mathcal{T}$ is the set of operators that are polynomials of the operator MD. Since each T_r commutes which each τ_a, so does any element of $\mathcal{T}$. These considerations naturally lead to the following assumption which will underpin the entire theory we develop:

Assumption 3: There exist $k \in \mathbb{N}$ and linear operators L, K such that X admits a Stein operator (in the sense of Definition 1.1) of the form

$$A = L - M^k K, \tag{6}$$

where the operators L, K are elements of $\mathcal{T}$.

In most situations that we consider, however, the operators K and L will be products of T_r operators. We now collect some useful relations for the operators that will be used throughout this paper.

Lemma 2.5. *Let $r, r' \in \mathbb{R}$, $a \in \mathbb{R} \setminus \{0\}$ and $n \in \mathbb{N}$. Then, acting on the class of functions $\mathcal{F} \supseteq C_0^\infty(\mathbb{R})$, the operators (M, D, T_r, τ_a) satisfy the following relations:*

$$\tau_a M = a\, M \tau_a \quad \textit{and} \quad D\tau_a = a\, \tau_a D,$$

and

$$T_r M^n = M^n T_{r+n} \quad \text{and} \quad T_r D^n = D^n T_{r-n}. \tag{7}$$

Additionally, T_r and $T_{r'}$ always commute, and every T_r commutes with every τ_a.

Proof. (i) Here, and throughout the proof, let $f \in \mathcal{F}$. Then, $\tau_a M f(x) = \tau_a x f(x) = axf(ax) = aM\tau_a f(x)$ and $D\tau_a f(x) = Df(ax) = af'(ax) = a\tau_a Df(x)$, as required.

(ii) By the product rule of differentiation, one has that $DM = MD + I$. Therefore, $T_r M = MDM + rM = M^2 D + (r+1)M = MT_{r+1}$, and the first relation now follows from direct recurrence. Similarly, we have $T_r D = rMD^2 + D = rDMD + (r-1)D = DT_{r-1}$, and the second relation now follows from direct recurrence.

(iii) That T_r and $T_{r'}$ commute follows since they are polynomials, of degree 1, in MD. Also, $T_r \tau_a f(x) = T_r f(ax) = axf'(ax) + rf(ax) = \tau_a(xf'(x) + rf(x)) = \tau_a T_r f(x)$, and therefore T_r and τ_a commute. □

Note also that if we define, for an operator L, $L\mathcal{T} : \{LA \, ; \, A \in \mathcal{T}\}$, and similarly $\mathcal{T}L$, then a direct consequence of (7) is that for any $n \in \mathbb{Z}_+$,

$$M^n \mathcal{T} = \mathcal{T} M^n \quad \text{and} \quad D^n \mathcal{T} = \mathcal{T} D^n. \tag{8}$$

2.3 More on Assumption 3

The purpose of this section is to give a simple criterion to identify operators A that satisfy Assumption 3. We have the following criterion for X to satisfy Assumption 3, when its Stein operator is written in an expanded form.

Lemma 2.6. *X satisfies Assumption 3 if, and only if, X has a Stein operator of the form $\sum_{i,j} a_{ij} M^j D^i$ with $\#\{j - i \,|\, a_{ij} \neq 0\} \leq 2$.*

Proof. Let us give a preliminary result. Let $L = M^{k_1} D^{l_1} \dots M^{k_n} D^{l_n}$. Let $q = \sum_i k_i - \sum_i l_i$. Then if $q \geq 0$, $L \in M^q \mathcal{T}$. Indeed, note that $MD = T_0$ and $DM = T_1$. Hence, in the product $M^{k_1} D^{l_1} \dots M^{k_n} D^{l_n}$, one can pair any product MD (or DM), replace it by T_0 (or T_1), and flush it right using (7). Hence the result.

Now we prove the Lemma. If X satisfies Assumption 3, then a Stein operator for X is $A = P(MD) - M^k Q(MD)$, where P and Q are polynomials. Using repeatedly the fact that $DM = MD + I$, one sees that for any integer $n \in \mathbb{N}$, $(MD)^n$ can be expanded in a sum of terms of the type $a_l M^l D^l$, $l \in \mathbb{N}$, $a_l \in \mathbb{R}$. The same holds for $Q(MD)$. Hence $A = \sum_{i,j} a_{ij} M^j D^i$ with $\{j - i \,|\, a_{ij} \neq 0\} = \{0, k\}$.

Let us prove the converse. We only treat the case $\#\{j - i \,|\, a_{ij} \neq 0\} = 2$, the others being similar. Assume $A = \sum_{i,j} a_{ij} M^j D^i$ such that $\#\{j - i \,|\, a_{ij} \neq 0\} = \{k_1, k_2\}$ with $k_1 > k_2$, is a Stein operator for X. Assume first that $k_2 \geq 0$. Then from the above, $M^j D^i \in M^{j-i}\mathcal{T}$, the latter set being either $M^{k_1}\mathcal{T}$ or $M^{k_2}\mathcal{T}$, depending on the value of $j - i$. But from (8), $M^{k_1}\mathcal{T} = M^{k_1 - k_2}\mathcal{T} M^{k_2}$ and $M^{k_2}\mathcal{T} = \mathcal{T} M^{k_2}$. Hence A can be rewritten $LM^{k_2} - M^{k_1 - k_2} K M^{k_2}$, with $L, K \in \mathcal{T}$. Simplifying by M^{k_2} on the right (i.e., applying A to $x \mapsto f(x)x^{-k_2}$ instead of f) yields the result.

If $k_2 < 0$, one can multiply A by M^{-k_2} on the right and proceed in the same manner. □

Remark 2.7. *Assume X admits a smooth density p, which solves the differential equation $Bp = 0$ with $B = \sum_{i,j} b_{ij} M^j D^i$. Then, by duality (i.e. integration by parts; see Section 4 for further detail), a Stein operator for X is given by $A = \sum_{i,j} (-1)^i b_{ij} D^i M^j$. Then, in a similar* *manner as in the previous lemma, one can prove that X satisfies Assumption 3 if, and only if, $\#\{j-i \,|\, b_{ij} \neq 0\} \leq 2$. In other words, the condition given in Lemma 2.6 for X to satisfy Assumption 3 can be equivalently checked on the Stein operator A or on the differential operator B which cancels out the density of X.*

One can specialize the result of Lemma 2.6 when the score of the distribution of X is a rational fraction, which includes a wide class of classical distributions. In this paper, X will be a continuous random variable, and we use the terminology score function of X to mean the logarithmic derivative of its probability density function.

Corollary 2.8. *Assume X admits a score function of the form*

$$\rho(x) := \frac{p'(x)}{p(x)} = \frac{a\,x^k + b\,x^l}{c\,x^{k+1} + d\,x^{l+1}}, \tag{9}$$

with $k, l \in \mathbb{N}$. Then X satisfies Assumption 3.

Conversely, if the score ρ of X is a rational fraction, and if X satisfies Assumption 3, then ρ is of the form (9)*.*

Proof. Assume that

$$\frac{p'(x)}{p(x)} = \frac{\sum_{i=0}^{n} a_i x^i}{\sum_{j=0}^{m} b_j x^j}.$$

Starting from the canonical Stein operator $f \mapsto f' + \frac{p'}{p} f$, and applying it to $(\sum_{j=0}^{m} b_j x^j) f(x)$, we have that $A = D\big(\sum_{j=0}^{n} b_j M^j\big) + \sum_{i=0}^{n} a_i M^i$ is a Stein operator for X. Since $DM^j = jM^{j-1} + M^j D$, an application of Lemma 2.6 yields the result. □

Let us now focus on the class of Pearson distributions, that is the collection all continuous probability distributions which satisfy 9 with $k = 1, l = 0$. We give here a generalization of [42], Theorem 1, p. 65 proving that Pearson distributions have polynomial Stein operators which can be written in terms of operators in $\mathcal{T}$.

Lemma 2.9 ([42],Theorem 1, p. 65)**.** *Let X be of Pearson type with score function*

$$\rho(x) := \frac{p'(x)}{p(x)} = -\frac{ax - \ell}{\delta_2 x^2 + \delta_1 x + \delta_0} \tag{10}$$

for some $a, \ell, \delta_0, \delta_1, \delta_2$ and x on the resulting support. If $\delta_0 = 0$ then

$$A = M\delta_2 T_{2-a/\delta_2} + \delta_1 T_{1+\ell/\delta_1} \tag{11}$$

and if $\delta_0 \neq 0$ then

$$A = M^2 \delta_2 T_{3-a/\delta_2} + \delta_1 M T_{1+\ell/\delta_1} + \delta_0 T_1 \tag{12}$$

is a Stein operator for X.

Proof. If X is Pearson with log-derivative (10) then denoting $-P_{\text{num}}/P_{\text{denom}}$ this ratio we see that $(f P_{\text{denom}} p)'/p = f(P'_{\text{denom}} - P_{\text{num}}) + f' P_{\text{denom}}$, i.e.

$$f'(x)\left(\delta_2 x^2 + \delta_1 x + \delta_0\right) + f(x)\left((2\delta_2 + a)x + (\delta_1 - \ell)\right).$$

This operator is integrable with respect to p (with integral 0) for all $f \in \mathcal{F}$. Conclusion (11) follows immediately, while (12) is obtained after replacing f with xf and using $T_r M = M T_{r+1}$. □

Example 2.10. *The normal distribution $N(\mu, \sigma^2)$ falls into this class. It has log-derivative $\rho(x) = -(x-\mu)/\sigma^2$ on $\mathbb{R}$: $a = 1$, $\ell = \mu$, $\delta_1 = \delta_2 = 0$ and $\delta_0 = \sigma^2$ and* (12) *applies leading to $A = \sigma^2 T_1 + \mu M - M^2$ (recall that $rT_{1/r} \to I$ as $r \to 0$). Assumption 3 is satisfied if and only if $\mu = 0$; in a future work [20] we shall introduce a technique for obtaining Stein operators for products of a class of distributions, which includes the non centered normal distribution, that do not satisfy Assumption 3. Other examples that are in the class of Lemma 2.9 include the gamma, beta, Student's t, and inverse-gamma distributions, all of which satisfy Assumption 3. The resulting Stein operators (which are not new to the literature) are given in Appendix A.*

2.4 The algebra for products of distributions

We first note how Proposition 2.4 is easily generalised to the product of n independent random variables, by induction. More precisely, if $(X_i)_{1\leq i\leq n}$ are independent random variables with respective Stein operator $L_i - M^p K_i$, if all the operators $\{L_i, K_i\}_{1\leq i\leq n}$ commute with each other and with the τ_a, $a \in \mathbb{R}$, then a Stein operator for $\prod_{i=1}^n X_i$ is

$$\prod_{i=1}^n L_i - M^p \prod_{i=1}^n K_i.$$

The main drawback of Proposition 2.4 is that we assume the same power of M^p appears in both operators. As such, the Proposition cannot be applied for instance for the product of a gamma (for which $p = 1$, see Appendix A) and a centered normal (for which $p = 2$, see Appendix A). In the following Lemma and Proposition, we show how to bypass this difficulty: one can build another Stein operator for X with the power p multiplied by an arbitrary integer k (even though by doing so, one increases the order of the operator). Here we restrict ourselves to the case where the L_i and K_i operators are products of operators T_α and we make use of the relation (7).

Lemma 2.11. *Assume X has a Stein operator of the form*

$$A_X = a\prod_{i=1}^n T_{\alpha_i} - bM^p \prod_{i=1}^m T_{\beta_i}. \tag{13}$$

Then, for every $k \geq 1$,

$$a^k \prod_{i=1}^n \prod_{j=0}^{k-1} T_{\alpha_i + jp} - b^k M^{kp} \prod_{i=1}^m \prod_{j=0}^{k-1} T_{\beta_i + jp}$$

is a Stein operator for X.

Proof. We prove the result by induction on k. By assumption, it is true for $k = 1$. Then, using the recurrence hypothesis and (7),

$$\begin{aligned}
\mathbb{E}\Big[a^{k+1} \prod_{j=0}^{k} \prod_{i=1}^n T_{\alpha_i + jk} f(X)\Big] &= \mathbb{E}\Big[a a^k \prod_{j=0}^{k-1} \prod_{i=1}^n T_{\alpha_i + jp} \Big(\prod_{i=1}^n T_{\alpha_i + kp} f\Big)(X)\Big] \\
&= \mathbb{E}\Big[a b^k M^{kp} \prod_{j=0}^{k-1} \prod_{i=1}^m T_{\beta_i + jp} \Big(\prod_{i=1}^n T_{\alpha_i + kp} f\Big)(X)\Big] \\
&= \mathbb{E}\Big[a b^k \prod_{i=1}^n T_{\alpha_i} M^{kp} \prod_{j=0}^{k-1} \prod_{i=1}^m T_{\beta_i + jp} f(X)\Big]
\end{aligned}$$

$$= \mathbb{E}\Big[b^{k+1}M^p \prod_{i=1}^m T_{\beta_i} M^{kp} \prod_{j=0}^{k-1}\prod_{i=1}^m T_{\beta_i+jp} f(X)\Big]$$
$$= \mathbb{E}\Big[b^{k+1}M^{kp} \prod_{j=0}^{k}\prod_{i=1}^m T_{\beta_i+jp} f(X)\Big],$$

which proves our claim. □

Lemma 2.11 leads to the following rule of thumb for the problem of finding a Stein operator for a product of independent random variables X and Y with Stein operators $A_X = a\prod_{i=1}^n T_{\alpha_i} - bM^p \prod_{i=1}^m T_{\beta_i}$ and $A_Y = a'\prod_{i=1}^{n'} T_{\alpha_i'} - b'M^{p'}\prod_{i=1}^{m'} T_{\beta_i'}$, with $p \neq p'$. Apply Lemma 2.11 to X with $k = p'$ and to Y with $k = p$ to get Stein operators for X and Y of the form of Proposition 2.4, but with p replaced by pp'. Then apply the Proposition. As an illustration, one can prove the following.

Proposition 2.12. *Assume X, Y are random variables with respective Stein operators*

$$\begin{aligned} A_X &= a_1 T_{\alpha_1} - a_2 M^p T_{\alpha_2}, \\ A_Y &= b_1 T_{\beta_1} - b_2 M^q T_{\beta_2}, \end{aligned}$$

where $p, q \in \mathbb{N}$ and $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R} \cup \{\infty\}$ (and Remark 2.3 applies in case any of the α_i or β_i, $i = 1, 2$ is set to $+\infty$). Let m be the least common multiple of p and q and write $m = k_1 p = k_2 q$. Then, if X and Y are independent,

$$a_1^{k_1} b_1^{k_2} \prod_{i=0}^{k_1-1} T_{\alpha_1+ip} \prod_{i=0}^{k_2-1} T_{\beta_1+iq} - M^m a_2^{k_1} b_2^{k_2} \prod_{i=0}^{k_1-1} T_{\alpha_2+ip} \prod_{i=0}^{k_2-1} T_{\beta_2+iq}$$

is a Stein operator for XY.

Proof. Apply Lemma 2.11 with k_1 and k_2 to get, for all $f \in \mathcal{F}$,

$$\mathbb{E}\Big[a_1^{k_1} \prod_{j=0}^{k_1-1} T_{\alpha_1+jp} f(X)\Big] = \mathbb{E}\Big[a_2^{k_1} M^m \prod_{j=0}^{k_1-1} T_{\alpha_2+jp} f(X)\Big],$$

and

$$\mathbb{E}\Big[b_1^{k_2} \prod_{j=0}^{k_2-1} T_{\beta_1+jp} f(Y)\Big] = \mathbb{E}\Big[b_2^{k_2} M^m \prod_{j=0}^{k_2-1} T_{\beta_2+jp} f(Y)\Big].$$

Then the proof follows from an application of Proposition 2.4. □

Remark 2.13. *Let us give an example when one of the T operators is the identity. We have that if X, Y are random variables with respective Stein operators*

$$\begin{aligned} A_X &= a_1 T_\alpha - a_2 M^p, \\ A_Y &= b_1 T_\beta - b_2 M^q, \end{aligned}$$

then, with the same notation, a Stein operator for XY is

$$a_1^{k_1} b_1^{k_2} \prod_{i=0}^{k_1-1} T_{\alpha+ip} \prod_{i=0}^{k_2-1} T_{\beta+iq} - a_2^{k_1} b_2^{k_2} M^m.$$

2.5 The algebra for powers and inverse distributions

In this section, we consider (not necessarily integer nor positive) powers of X's and this implies we shall need to modify Assumption 2 from Section 2.1 to ensure that all operators are well defined. If X takes values a.s. in $\mathbb{R}\backslash\{0\}$, then we restrict to test functions $f \in C_0^\infty(\mathbb{R}\backslash\{0\})$. If X takes values a.s. in $(0,\infty)$, then we further restrict to test functions $f \in C_0^\infty((0,\infty))$. Likewise, we extend some notations from Section 2.2. We extend the definition of M^a to $a \in \mathbb{Z}$ by $M^a f(x) = x^a f(x)$, $x \neq 0$. Further we extend this definition to $a \in \mathbb{R}$ for $x \in (0,\infty)$.

Let us first note a result concerning powers. Let P_a be defined by $P_a f(x) = f(x^a)$. For $a \neq 0$, we have that $T_r P_a = a P_a T_{r/a}$, since

$$\begin{aligned} T_r P_a f(x) &= x \cdot a x^{a-1} f'(x^a) + r f(x^a) = a x^a f'(x^a) + r f(x^a) \\ &= a(x^a f'(x^a) + (r/a) f(x^a)) = a P_a T_{r/a} f(x). \end{aligned}$$

This result allows us to easily obtain Stein operators for powers of random variables and inverse distributions. Suppose X has Stein operator

$$A_X = a T_{\alpha_1} \cdots T_{\alpha_n} - b M^q T_{\beta_1} \cdots T_{\beta_m}.$$

We can write down a Stein operator for X^γ immediately (if X takes negative values, we restrict to positive or integer-valued γ):

$$\begin{aligned} A_{X^\gamma} &= a T_{\alpha_1} \cdots T_{\alpha_n} P_\gamma - b M^q T_{\beta_1} \cdots T_{\beta_m} P_\gamma \\ &= a \gamma^n P_\gamma T_{\alpha_1/\gamma} \cdots T_{\alpha_n/\gamma} - b \gamma^m M^q P_\gamma T_{\beta_1/\gamma} \cdots T_{\beta_m/\gamma} \\ &= a \gamma^n P_\gamma T_{\alpha_1/\gamma} \cdots T_{\alpha_n/\gamma} - b \gamma^m P_\gamma M^{q/\gamma} T_{\beta_1/\gamma} \cdots T_{\beta_m/\gamma}. \end{aligned} \tag{14}$$

Applying $P_{1/\gamma}$ on the left of (14) gives the following Stein operator for the random variable X^γ:

$$\tilde{A}_{X^\gamma} = a \gamma^n T_{\alpha_1/\gamma} \cdots T_{\alpha_n/\gamma} - b \gamma^m M^{q/\gamma} T_{\beta_1/\gamma} \cdots T_{\beta_m/\gamma}, \tag{15}$$

as $P_{1/\gamma} P_\gamma = I$.

From (15) we immediately obtain, for example, the classical $\chi^2_{(1)}$ Stein operator $T_{1/2} - \frac{1}{2}M$ from the standard normal Stein operator $T_1 - M^2$. However, in certain situations, a more convenient form of the Stein operator may be desired. To illustrate this, we consider the important special case of inverse distributions. Here $\gamma = -1$, which yields the following Stein operator for $1/X$:

$$a(-1)^n T_{-\alpha_1} \cdots T_{-\alpha_n} - b(-1)^m M^{-q} T_{-\beta_1} \cdots T_{-\beta_m}.$$

To remove the singularity, we multiply on the right by M^{-1} to get

$$\begin{aligned} A_{1/X} &= a(-1)^n T_{-\alpha_1} \cdots T_{-\alpha_n} M^q - b(-1)^m M^{-q} T_{-\beta_1} \cdots T_{-\beta_m} M^q \\ &= a(-1)^n M^q T_{q-\alpha_1} \cdots T_{q-\alpha_n} - b(-1)^m T_{q-\beta_1} \cdots T_{q-\beta_m}. \end{aligned}$$

Cancelling constants gives the Stein operator

$$\tilde{A}_{1/X} = b T_{q-\beta_1} \cdots T_{q-\beta_m} - (-1)^{m+n} a M^q T_{q-\alpha_1} \cdots T_{q-\alpha_n}. \tag{16}$$

3 Applying the algebra to find new Stein operators

Starting from the classical Stein operators of the centered normal, gamma, beta, Student's t, inverse-gamma, PRR, variance-gamma (with $\theta = 0$ and $\mu = 0$), and generalized gamma distributions, we use the results of Section 2.4 to derive new operators for the (possibly mixed) products of these distributions. The operators of the aforementioned distributions are summed up in Appendix A. Stein operators for any mixed product of independent copies of such random variables are attainable through a direct application of Proposition 2.12. We give some examples below.

3.1 Mixed products of centered normal and gamma random variables

Stein operators for (mixed) products of independent central normal, beta and gamma random variables were obtained by [16, 19]. Here we demonstrate how these Stein operators can be easily derived by an application of our theory (we omit the beta distribution for reasons of brevity). Let $(X_i)_{1\leq i\leq n}$ and $(Y_j)_{1\leq j\leq m}$ be independent random variables and assume $X_i \sim \mathcal{N}(0,\sigma^2)$ and $Y_j \sim \Gamma(r_j, \lambda_j)$. The random variables X_i and Y_j admit the following Stein operators:

$$\begin{aligned} A_{X_i} &= \sigma_i^2 T_1 - M^2, && (17)\\ A_{Y_j} &= T_{r_j} - \lambda_j M. && (18) \end{aligned}$$

A repeated application of Proposition 2.12 now gives the following Stein operators:

$$\begin{aligned} A_{X_1\cdots X_n} &= \sigma_1^2\cdots\sigma_n^2 T_1^n - M^2, && (19)\\ A_{Y_1\cdots Y_m} &= T_{r_1}\cdots T_{r_m} - \lambda_1\cdots\lambda_m M, && (20)\\ A_{X_1\cdots X_n Y_1\cdots Y_m} &= \sigma_1^2\cdots\sigma_n^2 T_1^n T_{r_1}\cdots T_{r_m} T_{r_1+1}\cdots T_{r_m+1} - \lambda_1\cdots\lambda_m M^2. && (21) \end{aligned}$$

The product gamma Stein operator (20) is in exact agreement with the one obtained by [19]. However, the Stein operators (19) and (21) differ slightly from those of [16, 19], because they act on different functions. Indeed, the product normal Stein operator given in [16] is $\tilde{A}_{X_1\cdots X_n} = \sigma_1^2\cdots\sigma_n^2 D T_0^n - M$, but multiplying through on the right by M yields (19). The same is true of the mixed product operator (21), which is equivalent to the mixed normal-gamma Stein operator of [19] multiplied on the right by M. We refer to Appendix A where this idea is expounded.

Finally, we note that whilst the operators (19) and (20) are of orders n and m, respectively, the mixed product operator (21) is of order $n + 2m$, rather than order $n + m$ which one may at first expect. This a consequence of the fact that the powers of M in the Stein operator (17) and (18) differ by a factor of 2.

3.2 Mixed product of Student and variance-gamma random variables

Let $(X_i)_{1\leq i\leq n}$ and $(Y_j)_{1\leq j\leq m}$ be independent random variables and assume $X_i \sim t_{\nu_i}$ follows Student's t-distribution with ν degrees of freedom and $Y_j \sim \mathrm{VG}(r_j, 0, \sigma_j, 0)$; the p.d.f.s of these distributions are given in Appendix A. X_i and Y_j admit Stein operators of the form:

$$\begin{aligned} A_{X_i} &= \nu_i T_1 + M^2 T_{2-\nu_i},\\ A_{Y_j} &= \sigma_j^2 T_1 T_{r_j} - M^2. && (22) \end{aligned}$$

Note that one cannot apply Proposition 2.4 to the $\mathrm{VG}(r,\theta,\sigma,0)$ Stein operator $\sigma^2 T_1 T_r + 2\theta M T_{r/2} - M^2$, because Assumption 3 is not satisfied.

Applying recursively Proposition 2.4, we obtain the following Stein operators:

$$\begin{aligned} A_{X_1\cdots X_n} &= \nu_1\ldots\nu_n T_1^n-(-1)^n M^2 T_{2-\nu_1}\ldots T_{2-\nu_n}, \\ A_{Y_1\cdots Y_m} &= \sigma_1^2\ldots\sigma_m^2 T_1^m T_{r_1}\ldots T_{r_m}-M^2, \\ A_{X_1\cdots X_n Y_1\cdots Y_m} &= \nu_1\ldots\nu_n\sigma_1^2\ldots\sigma_m^2 T_1^{n+m}T_{r_1}\ldots T_{r_m}-(-1)^n M^2 T_{2-\nu_1}\ldots T_{2-\nu_n}. \end{aligned} \tag{23}$$

As an aside, note that (22) can be obtained by applying Proposition 2.4 to the Stein operators

$$A_X=\sigma^2 T_1-M^2, \quad A_Y=T_r-M^2,$$

where X and Y are independent. We can identify A_X as the Stein operator for a $\mathcal{N}(0,\sigma^2)$ random variable and A_Y as the Stein operator of the random variable $Y=\sqrt{V}$ where $V\sim\Gamma(r/2,1/2)$. Since the variance-gamma Stein operator is characterizing (see [15], Lemma 3.1), it follows that that $Z\sim\mathrm{VG}(r,0,\sigma,0)$ is equal in distribution to $X\sqrt{V}$. This representation of the $\mathrm{VG}(r,0,\sigma,0)$ distribution can be found in [6]. This example demonstrates that by characterizing probability distributions, Stein operators can be used to derive useful properties of probability distributions; for a further discussion on this general matter see Section 4.

The Stein operator (23) will be used in Section 4 in a derivation of the formula for the p.d.f. of the product of independent $\mathrm{VG}(r,0,\sigma,0)$ random variables. As an example, following some straightforward calculations, we write down explicitly the Stein operator for the case $m=2$ and $\sigma_1=\sigma_2=1$:

$$\begin{aligned} A_{Y_1Y_2}f(x) &= x^4f^{(4)}(x)+(8+r_1+r_2)x^3f^{(3)}(x)+(17+5r_1+5r_2+r_1r_2)x^2f''(x) \\ &\quad+(4+3r_1+3r_2+r_1r_2)xf'(x)+(r_1r_2-x^2)f(x). \end{aligned}$$

3.3 PRR distribution

A Stein operator for the PRR distribution is given by

$$A_sf(x)=sT_1T_2f(x)-M^2T_{2s}f(x)=sx^2f''(x)+(4sx-x^3)f'(x)+2s(1-x^2)f(x), \tag{24}$$

see Appendix A. We now exhibit a neat derivation of this Stein operator by an application of Sections 2.4 and 2.5. Let X and Y be independent random variables with distributions

$$X\sim\begin{cases}\mathrm{Beta}(1,s-1), & \text{if } s>1,\\ \mathrm{Beta}(1/2,s-1/2), & \text{if } 1/2<s\le 1,\end{cases}$$

and

$$Y\sim\begin{cases}\Gamma(1/2,1), & \text{if } s>1,\\ \mathrm{Exp}(1), & \text{if } 1/2<s\le 1.\end{cases}$$

Then it is known that $\sqrt{2sXY}\sim K_s$ (see [34], Proposition 2.3).

If $s>1$, then we have the following Stein operators for X and Y:

$$A_X=T_1-MT_s, \quad A_Y=T_{1/2}-M,$$

and, for $1/2<s\le 1$, we have the following Stein operators for X and Y:

$$A_X=T_{1/2}-MT_s, \quad A_Y=T_1-M.$$

Using Proposition 2.12, we have that, for all $s > 1/2$,

$$A_{XY} = T_{1/2}T_1 - MT_s.$$

From (15) we obtain the Stein operator

$$A_{\sqrt{XY}} = T_1T_2 - 2M^2T_{2s},$$

which on rescaling by a factor of $\sqrt{2s}$ yields the operator (24).

3.4 Inverse and quotient distributions

From (16) we can write down inverse distributions for many standard distributions. First, suppose $X \sim \mathrm{Beta}(a, b)$. Then a Stein operator for $1/X$ is

$$A_{1/X} = T_{1-a-b} - MT_{1-a}. \tag{25}$$

Now, let $X_1 \sim \mathrm{Beta}(a_1, b_1)$ and $X_2 \sim \mathrm{Beta}(a_2, b_2)$ be independent. Then using Proposition 2.12 applied to the Stein operator (25) for $1/X$ and the beta Stein operator, we have the following Stein operator for $Z = X_1/X_2$:

$$A_Z = T_{a_1}T_{1-a_2-b_2} - MT_{a_1+b_1}T_{1-a_2}, \tag{26}$$

which is a second order differential operator.

Let us consider the inverse-gamma distribution. Let $X \sim \Gamma(r, \lambda)$, then the gamma Stein equation is $A_X = T_r - \lambda M$. From (16) we can obtain a Stein operator for $1/X$ (an inverse-gamma random variable):

$$A_{1/X} = MT_{1-r} - \lambda I.$$

If $X_1 \sim \Gamma(r_1, \lambda_1)$ and $X \sim \Gamma(r_2, \lambda_2)$, we have from the above operator and Proposition 2.12, the following Stein operator for $Z = X_1/X_2$:

$$A_Z = \lambda_1 MT_{1-r_2} + \lambda_2 T_{r_1}, \tag{27}$$

which is a first order differential operator. As a special case, we can obtain a Stein operator for the F-distribution with parameters $d_1 > 0$ and $d_2 > 0$. This is because $Z \sim F(d_1, d_2)$ is equal in distribution to $\frac{X_1/d_1}{X_2/d_2}$, where $X_1 \sim \chi^2_{(d_1)}$ and $X_2 \sim \chi^2_{(d_2)}$ are independent. Now applying (27) and rescaling to take into account the factor d_1/d_2 gives the following Stein operator for Z:

$$A_Z = d_1 MT_{1-d_2/2} + d_2 T_{d_1/2}. \tag{28}$$

As this Stein operator seems to be new to the literature and may prove useful in applications due to the importance of the F-distribution in statistics, we write out the operator explicitly:

$$A_Z f(x) = (d_1x^2 + d_2x)f'(x) + (d_1d_2/2 + d_1(1 - d_2/2)x)f(x).$$

One can also easily derive the generalized gamma Stein operator from the gamma Stein operator. The Stein operator for the $\mathrm{GG}(r, \lambda, q)$ distribution is given by $T_r - q\lambda^q M^q$. Using the relationship $X \stackrel{\mathcal{L}}{=} (\lambda^{1-q}Y)^{1/q}$ for $X \sim \mathrm{GG}(r, \lambda, q)$ and $Y \sim \Gamma(r/q, \lambda)$ (see [35]) together with (15) and a rescaling, we readily recover the generalized gamma Stein operator from the usual gamma Stein operator.

As a final example, we note that we can use Proposition 2.12 to obtain a Stein operator for the ratio of two independent standard normal random variables. A Stein operator for the standard normal random variable X_1 is $T_1 - M^2$ and we can apply (16) to obtain the following Stein operator for the random variable $1/X_1$:

$$A_{1/X_1} = M^2 T_1 - I.$$

Hence a Stein operator for the ratio of two independent standard normals is

$$A = (1 + M^2)T_1,$$

which is the Stein operator for the Cauchy distribution (a special case of the Student's t Stein operator of [38]), as one would expect.

4 Duals of Stein operators and densities of product distributions

Fundamental methods, based on the Mellin integral transform, for deriving formulas for densities of product distributions were developed by [39, 40]. In [40], formulas, involving the Meijer G-function, were obtained for products of independent centered normals, and for mixed products of beta and gamma random variables. However, for other product distributions, applying the Mellin inversion formula can lead to intractable calculations.

In this section, we present a novel method for deriving formulas for densities of product distributions based on the duality between Stein operators and ODEs satisfied by densities. Our approach builds on that of [19] in which a duality argument was used to derive a new formula for the p.d.f. of a mixed product of mutually independently centered normal, beta and gamma random variables (deriving such a formula using the Mellin inversion formula would have required some very involved calculations). We apply this method to derive a new formula for the p.d.f. of the product of n independent $\mathrm{VG}(r, 0, \sigma, 0)$ random variables. We begin with a duality lemma whose proof is a generalisation of the argument given in Section 3.2 of [19].

Lemma 4.1. *Let Z be a random variable with density p supported on an interval $[a, b] \subseteq \mathbb{R}$. Let*

$$Af(x) = T_{r_1} \cdots T_{r_n} f(x) - bx^q T_{a_1} \cdots T_{a_m} f(x), \tag{29}$$

and suppose that

$$\mathbb{E}[Af(Z)] = 0 \tag{30}$$

for all $f \in C^k([a, b])$, where $k = \max\{m, n\}$, such that

1. *$x^{q+1+i+j} p^{(i)}(x) f^{(j)}(x) \to 0$, as $x \to a$ and as $x \to b$, for all i, j such that $0 \le i + j \le m$;*
2. *$x^{1+i+j} p^{(i)}(x) f^{(j)}(x) \to 0$, as $x \to a$ and as $x \to b$, for all i, j such that $0 \le i + j \le n$.*

(We denote this class of functions by $\mathcal{C}_p$). Then p satisfies the differential equation

$$T_{1-r_1} \cdots T_{1-r_n} p(x) - b(-1)^{m+n} x^q T_{q+1-a_1} \cdots T_{q+1-a_m} p(x) = 0. \tag{31}$$

Remark 4.2. *The class of functions $\mathcal{C}_p$ consists of all $f \in C^k([a, b])$, where $k = \max\{m, n\}$, that satisfy particular boundary conditions at a and b. Note that when $(a, b) = \mathbb{R}$ the class includes the set of all functions on $\mathbb{R}$ with compact support that are k times differentiable. The class $\mathcal{C}_p$ suffices for the purpose of deriving the differential equation (31), although we expect that for particular densities (such as the beta distribution) the conditions on f could be weakened.*

Proof. We begin by writing the expectation (30) as

$$\int_a^b \big\{T_{r_1}\cdots T_{r_n}f(x) - bx^q T_{a_1}\cdots T_{a_m}f(x)\big\}p(x)\,\mathrm{d}x = 0, \tag{32}$$

which exists if $f \in \mathcal{C}_p$. In arriving at the differential equation (31), we shall apply integration by parts repeatedly. To this end, it is useful to note the following integration by parts formula. Let $\gamma \in \mathbb{R}$ and suppose that ϕ and ψ are differentiable. Then

$$\begin{aligned}\int_a^b x^\gamma \phi(x) T_r\psi(x)\,\mathrm{d}x &= \int_a^b x^\gamma \phi(x)\{x\psi'(x) + r\psi(x)\}\,\mathrm{d}x = \int_a^b x^{\gamma+1-r}\phi(x)\frac{\mathrm{d}}{\mathrm{d}x}(x^r\psi(x))\,\mathrm{d}x \\ &= \Big[x^{\gamma+1}\phi(x)\psi(x)\Big]_a^b - \int_a^b x^r\psi(x)\frac{\mathrm{d}}{\mathrm{d}x}(x^{\gamma+1-r}\phi(x))\,\mathrm{d}x \\ &= \Big[x^{\gamma+1}\phi(x)\psi(x)\Big]_a^b - \int_a^b x^\gamma\psi(x)T_{\gamma+1-r}\phi(x)\,\mathrm{d}x, \end{aligned} \tag{33}$$

provided the integrals exist.

We now return to equation (32) and use the integration by parts and formula (33) to obtain a differential equation that is satisfied by p. Using (33) we obtain

$$\begin{aligned}\int_a^b x^q p(x) T_{a_1}\cdots T_{a_m}f(x)\,\mathrm{d}x &= \Big[x^{q+1}p(x)T_{a_2}\cdots T_{a_m}f(x)\Big]_a^b \\ &\quad - \int_a^b x^q T_{q+1-a_1}p(x)T_{a_2}\cdots T_{a_m}f(x)\,\mathrm{d}x \\ &= -\int_a^b x^q T_{q+1-a_1}p(x)T_{a_2}\cdots T_{a_m}f(x)\,\mathrm{d}x,\end{aligned}$$

where we used condition (i) to obtain the last equality. By a repeated application of integration by parts, using formula (33) and condition (i), we arrive at

$$\int_a^b x^q p(x)T_{a_1}\cdots T_{a_m}f(x)\,\mathrm{d}x = (-1)^m\int_a^b x^q f(x)T_{q+1-a_1}\cdots T_{q+1-a_m}p(x)\,\mathrm{d}x.$$

By a similar argument, this time using formula (33) and condition (ii), we obtain

$$\int_a^b p(x)T_{r_1}\cdots T_{r_n}f(x)\,\mathrm{d}x = (-1)^n\int_a^b f(x)T_{1-r_1}\cdots T_{1-r_n}p(x)\,\mathrm{d}x.$$

Putting this together we have that

$$\int_a^b \{(-1)^n T_{1-r_1}\cdots T_{1-r_n}p(x) - b(-1)^m x^q T_{q+1-a_1}\cdots T_{q+1-a_m}p(x)\}f(x)\,\mathrm{d}x = 0 \tag{34}$$

for all $f \in \mathcal{C}_p$. Since (34) holds for all $f \in \mathcal{C}_p$, we deduce (from an argument analogous to that used to prove the fundamental lemma of the calculus of variations) that p satisfies the differential equation (31). This completes the proof. $\square$

We now show how the duality Lemma 4.1 can be exploited to derive formulas for densities of distributions. By duality, p satisfies the differential equation (31), and making the change of variables $y = \frac{b}{q^{n-m}}x^q$ yields the following differential equation

$$T_{\frac{1-r_1}{q}}\cdots T_{\frac{1-r_n}{q}}p(y) - (-1)^{m+n}yT_{\frac{q+1-a_1}{q}}\cdots T_{\frac{q+1-a_m}{q}}p(y) = 0. \tag{35}$$

We recognise (35) as an instance of the Meijer G-function differential equation (41). There are $\max\{m,n\}$ linearly independent solutions to (35) that can be written in terms of the Meijer G-function (see [33], Chapter 16, Section 21). Using a change of variables, we can thus obtain a fundamental system of solutions to (31) given as Meijer G-functions. One can then arrive at a formula for the density by imposing the conditions that the solution must be non-negative and integrate to 1 over the support of the distribution. Due to the difficulty of handling the Meijer G-function, this final analysis is in general not straightforward. However, one can "guess" a formula for the density based on the fundamental system of solutions, and then verify that this is indeed the density by an application of the Mellin transform (note that in this verification step there is no need to use the Mellin inversion formula). An interesting direction for future research would be to develop techniques for identifying formulas for densities of distributions based solely on an analysis of the differential equation (31). However, even as it stands, we have a technique for obtaining formulas for densities that may be intractable through standard methods.

Products of $\mathrm{VG}(r,0,\sigma,0)$ *random variables.* Let $(Z_i)_{1\leq i\leq n}\sim\mathrm{VG}(r_i,0,\sigma_i,0)$ be independent, and set $Z=\prod_{i=1}^n Z_i$. Recall the Stein operator (23) for the product of $\mathrm{VG}(r_i,0,\sigma_i,0)$ distributed random variables:

$$A_Zf(x)=\sigma^2T_1^nT_{r_1}\dots T_{r_n}-M^2,$$

where $\sigma^2=\sigma_1^2\dots\sigma_n^2$. By Lemma 4.1, it follows that the density p satisfies the following differential equation:

$$T_0^nT_{1-r_1}\cdots T_{1-r_n}p(x)-\sigma^{-2}x^2p(x)=0. \tag{36}$$

Arguing as we did to obtain the ODE (35), we make the substitution $y=\frac{x^2}{2^{2n}\sigma^2}$ to reduce (36) to a G-function ODE of the type (41). We can therefore identify that the following function is a solution to (36):

$$p(x)=CG^{2n,0}_{0,2n}\bigg(\frac{x^2}{2^{2n}\sigma^2}\;\bigg|\;\frac{r_1-1}{2},\dots,\frac{r_n-1}{2},0,\dots,0\bigg),$$

where C is an arbitrary constant. We can apply (40) to choose C such the p integrates to 1 across its support:

$$p(x)=\frac{1}{2^n\pi^{n/2}\sigma}\prod_{j=1}^n\frac{1}{\Gamma(r_j/2)}G^{2n,0}_{0,2n}\bigg(\frac{x^2}{2^{2n}\sigma^2}\;\bigg|\;\frac{r_1-1}{2},\dots,\frac{r_n-1}{2},0,\dots,0\bigg). \tag{37}$$

We verify that this 'guess' this is indeed the density using Mellin transforms; note that this verification is much more straightforward than an application of the Mellin inversion formula.

Let us define the Mellin transform and state some properties that will be useful to us. The Mellin transform of a non-negative random variable U with density p is given by $M_U(s)=\mathbb{E}U^{s-1}$, for all s such that the expectation exists. If the random variable U has density p that is symmetric about the origin then we can define the Mellin transform of U by

$$M_U(s)=2\int_0^\infty x^{s-1}p(x)\,\mathrm{d}x. \tag{38}$$

The Mellin transform is useful in determining the distribution of products of independent random variables due to the property that if the random variables U and V are independent then $M_{UV}(s)=M_U(s)M_V(s)$.

To obtain the Mellin transform of $Z=\prod_{i=1}^n Z_i$, we recall that $Z_i\stackrel{\mathcal{L}}{=}X_i\sqrt{Y_i}$, where $X_i\sim\mathcal{N}(0,\sigma_i^2)$ and $Y_i\sim\Gamma(r/2,1/2)$ are independent. Using the formulas for the Mellin transforms of

Distribution	Parameters	Notation
Normal	$\mu, \sigma \in \mathbb{R}$	$\mathcal{N}(\mu, \sigma^2)$
Gamma	$r, \lambda > 0$	$\Gamma(r, \lambda)$
Beta	$a, b > 0$	$\mathrm{Beta}(a, b)$
Student's t	$\nu > 0$	t_ν
Inverse-gamma	$\alpha, \beta > 0$	$IG(\alpha, \beta)$
F-distribution	$d_1, d_2 > 0$	$F(d_1, d_2)$
PRR distribution	$s > 1/2$	PRR_s
Variance-gamma	$r, \sigma > 0,\ \theta, \mu \in \mathbb{R}$	$\mathrm{VG}(r, \theta, \sigma, \mu)$
Generalized gamma	$r, \lambda, q > 0$	$\mathrm{GG}(r, \lambda, q)$

Table 1: Distributions

the normal and gamma distributions (see [40]), we have that

$$M_{X_i}(s) = \frac{1}{\sqrt{\pi}} 2^{(s-1)/2} \sigma_i^{s-1} \Gamma(1/2), \quad M_{\sqrt{Y_i}}(s) = M_{Y_i}((s+1)/2) = 2^{(s-1)/2} \frac{\Gamma(\frac{r_i - 1 + s}{2})}{\Gamma(r_i)},$$

and therefore

$$M_Z(s) = \frac{1}{\pi^{n/2}} 2^{n(s-1)} \sigma^{s-1} [\Gamma(s/2)]^n \prod_{i=1}^n \frac{\Gamma(\frac{r_i - 1 + s}{2})}{\Gamma(\frac{r_i}{2})}. \tag{39}$$

Now, let W denote a random variable with density (37). Then, using (38) and (40) gives that

$$M_W(s) = 2 \times \frac{1}{2^n \pi^{n/2} \sigma} \prod_{j=1}^n \frac{1}{\Gamma(r_j/2)} \times \left(\frac{1}{2^{2n}\sigma^2}\right)^{-s/2} \times [\Gamma(s/2)]^n \times \prod_{i=1}^n \Gamma\left(\frac{r_i - 1 + s}{2}\right),$$

which is equal to (39). Since the Mellin transforms of W is equal to that of Z, it follows that W is equal in law to Z. Therefore (37) is indeed the p.d.f. of the random variable Z.

A List of Stein operators for continuous distributions

Recall that $Mf(x) = xf(x)$, I is the identity and $T_a f(x) = xf'(x) + af(x)$. We also recall that the beta function is defined by $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, and that $U(a, b, x)$ and $K_\nu(x)$ denote the confluent hypergeometric function of the second kind ([33], Chapter 13) and the modified Bessel function of the second kind ([33], Chapter 10), respectively.

We give a list of Stein operators for several classical probability distributions, in terms of the above operators. References for these Stein operators are as follows: normal [41], gamma [10, 27], beta [11, 21, 38], Student's t [38], inverse-gamma [23], F-distribution (new to this paper), PRR [34], variance-gamma [15], and generalized gamma [19].

The usual Stein operators (as defined in the above references) for the normal, PRR and variance-gamma distributions, are not in the form required in Section 2. In these cases, we multiply the operators by M on the right for the normal and variance-gamma distributions, and we multiply the operator by M^2 on the right for the PRR distribution.

Distribution	p.d.f.	Stein operator
$\mathcal{N}(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-\mu)/\sigma^2}$	$\sigma^2T_1+\mu M-M^2$
$\Gamma(r,\lambda)$	$\frac{\lambda^r}{\Gamma(r)}x^{r-1}\mathrm{e}^{-\lambda x}\mathbf{1}_{x>0}$	$T_r-\lambda M.$
$\mathrm{Beta}(a,b)$	$\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}\mathbf{1}_{0<x<1}$	T_a-MT_{a+b}
t_ν	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\big(1+\frac{x^2}{\nu}\big)^{-(\nu+1)/2}$	$\nu T_1+M^2T_{2-\nu}$
$IG(\alpha,\beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}\mathrm{e}^{-\beta/x}\,\mathbf{1}_{x>0}$	$\beta I+MT_{1-\alpha}$
$F(d_1,d_2)$	$\frac{1}{B(\frac{d_1}{2},\frac{d_2}{2})}\big(\frac{d_1}{d_2}\big)^{d_1/2}x^{d_1/2-1}\big(1+\frac{d_1}{d_2}x\big)^{-(d_1+2_2)/2}\,\mathbf{1}_{x>0}$	$d_2T_{d_1/2}+d_1MT_{1-d_2/2}$
PRR_s	$\Gamma(s)\sqrt{\frac{2}{s\pi}}\exp\big(-\frac{x^2}{2s}\big)U\big(s-1,\frac{1}{2},\frac{x^2}{2s}\big)\mathbf{1}_{x>0}$	$sT_1T_2-M^2T_{2s}$
$\mathrm{VG}(r,\theta,\sigma,\mu=0)$	$\frac{1}{\sigma\sqrt{\pi}\Gamma(\frac{r}{2})}\mathrm{e}^{\frac{\theta}{\sigma^2}x}\big(\frac{\lvert x\rvert}{2\sqrt{\theta^2+\sigma^2}}\big)^{\frac{r-1}{2}}K_{\frac{r-1}{2}}\big(\frac{\sqrt{\theta^2+\sigma^2}}{\sigma^2}\lvert x\rvert\big)$	$\sigma^2T_1T_r+2\theta MT_{r/2}-M^2$
$\mathrm{GG}(r,\lambda,q)$	$\frac{q\lambda^r}{\Gamma(r/q)}x^{r-1}\mathrm{e}^{-(\lambda x)^q}\,\mathbf{1}_{x>0}$	$T_r-q\lambda^qM^q$

Table 2: p.d.f. and Stein operator of some classical distributions.

B The Meijer G-function

The Meijer G-function is defined (see see [28, 33]), for $z\in\mathbb{C}\setminus\{0\}$, by the contour integral:

$$G^{m,n}_{p,q}\left(z\;\middle|\;\begin{matrix}a_1,\dots,a_p\\ b_1,\dots,b_q\end{matrix}\right)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}z^{-s}\frac{\prod_{j=1}^m\Gamma(s+b_j)\prod_{j=1}^n\Gamma(1-a_j-s)}{\prod_{j=n+1}^p\Gamma(s+a_j)\prod_{j=m+1}^q\Gamma(1-b_j-s)}\,\mathrm{d}s,$$

where c is a real constant defining a Bromwich path separating the poles of $\Gamma(s+b_j)$ from those of $\Gamma(1-a_j-s)$ and where we use the convention that the empty product is 1.

The following formula follows from [28], formula (1) of Section 5.6 and a change of variables:

$$\int_0^\infty x^{s-1}G^{m,n}_{p,q}\left(\alpha x^\gamma\;\middle|\;\begin{matrix}a_1,\dots,a_p\\ b_1,\dots,b_q\end{matrix}\right)\mathrm{d}x=\frac{\alpha^{-s/\gamma}}{\gamma}\frac{\prod_{j=1}^m\Gamma(b_j+\frac{s}{\gamma})\prod_{j=1}^n\Gamma(1-a_j-\frac{s}{\gamma})}{\prod_{j=m+1}^q\Gamma(1-b_j-\frac{s}{\gamma})\prod_{j=n+1}^p\Gamma(a_j+\frac{s}{\gamma})}.\quad(40)$$

For the conditions under which this formula is valid see [28], pp. 158–159. In particular, the formula is valid when $n=0$, $1\le p+1\le m\le q$ and $\alpha>0$.

The G-function $f(z)=G^{m,n}_{p,q}\big(z\,\big|\,{a_1,\dots,a_p \atop b_1,\dots,b_q}\big)$ satisfies the differential equation

$$(-1)^{p-m-n}zT_{1-a_1}\cdots T_{1-a_p}f(z)-T_{-b_1}\cdots T_{-b_q}f(z)=0.\quad(41)$$

Acknowledgements

The authors would like to thank the referees for their constructive comments, particularly to one referee for their careful reading of our paper and their suggestions which lead to a substantial improvement in the organisation of the paper. RG acknowledges support from EPSRC grant EP/K032402/1 and is currently supported by a Dame Kathleen Ollerenshaw Research Fellowship. RG is grateful to Université de Liège, FNRS and EPSRC for funding a visit to University de Liège, where some of the details of this project were worked out. YS gratefully acknowledges support by the Fonds de la Recherche Scientifique - FNRS under Grant MIS F.4539.16. Part of GM's research was supported by a WG (Welcome Grant) from Université de Liège.

References

[1] Arras, B., Azmoodeh, E., Poly, G. and Swan, Y. Stein characterizations for linear combinations of gamma random variables. arXiv:1709.01161, 2017.

[2] Arras, B., Azmoodeh, E., Poly, G. and Swan, Y. A bound on the 2-Wasserstein distance between linear combinations of independent random variables. To appear in *Stoch. Proc. Appl.*, 2018+.

[3] Arras, B. and Houdré, C. On Stein's method for infinitely divisible laws with finite first moment. arXiv:1712.10051, 2017.

[4] Arras, B., Mijoule, G., Poly, G. and Swan, Y. A new approach to the Stein-Tikhomirov method, with applications to the second Wiener chaos and Dickman convergence. arXiv:1605.06819, 2017.

[5] Barbour, A. D., Holst, L. and Janson, S. *Poisson Approximation*. Oxford University Press, Oxford, 1992.

[6] Barndorff-Nielsen, O. E., Kent, J. and Sørensen, M. Normal Variance-Mean Mixtures and z Distributions. *Int. Stat. Rev.* **50** (1982), pp. 145–159.

[7] Chatterjee, S., Fulman, J. and Röllin, A. Exponential approximation by Stein's method and spectral graph theory. *ALEA Lat. Am. J. Probab. Math. Stat.* **8** (2011) pp. 197-223.

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

[9] Chen, L. H. Y., Goldstein, L. and Shao, Q. M. *Normal approximation by Stein's method*. Springer Science & Business Media, 2010.

[10] Diaconis, P. and Zabell, S. Closed Form Summation for Classical Distributions: Variations on a Theme of De Moivre. *Statist. Sci.* **6** (1991), pp. 284–302.

[11] Döbler, C. Stein's method of exchangeable pairs for the beta distribution and generalizations. *Electron. J. Probab.* **20** no. 109 (2015), pp. 1–34.

[12] Döbler, C., Gaunt, R. E. and Vollmer, S. J. An iterative technique for bounding derivatives of solutions of Stein equations. *Electron. J. Probab.* **22** no. 96 (2017), pp. 1–39.

[13] Eichelsbacher and P. Löwe, M. Stein's Method for Dependent Random Variables Occurring in Statistical Mechanics. *Electron. J. Probab.* **15** (2010), pp. 962–988.

[14] Eichelsbacher, P. and Thäle, C. Malliavin-Stein method for Variance-Gamma approximation on Wiener space. *Electron. J. Probab.* **20** no. 123 (2015), pp. 1–28.

[15] Gaunt, R. E. Variance-Gamma approximation via Stein's method. *Electron. J. Probab.* **19** no. 38 (2014), pp. 1–33.

[16] Gaunt, R. E. On Stein's method for products of normal random variables and zero bias couplings. *Bernoulli* **23** (2017), pp. 3311–3345.

[17] Gaunt, R. E. A Stein characterisation of the generalized hyperbolic distribution. *ESAIM: PS* **21** (2017), pp. 303–316.

[18] Gaunt, R. E. A probabilistic proof of some integral formulas involving the Meijer G-function. *Ramanujan J.* **45** (2018), pp. 253–264.

[19] Gaunt, R. E. Products of normal, beta and gamma random variables: Stein operators and distributional theory. *Braz. J. Probab. Stat.* **32** (2018), pp. 437–466.

[20] Gaunt, R. E., Miljoule, G. and Swan, Y. Some new Stein operators. *In preparation*, 2018+.

[21] Goldstein, L. and Reinert, G. Stein's method for the Beta distribution and the Pólya-Eggenberger Urn. *J. Appl. Probab.* **50** (2013), pp. 1187–1205.

[22] Harper, A. J. Two new proofs of the Erdös–Kac theorem, with bound on the rate of convergence, by Stein's method for distributional approximations. *Math. Proc. Camb. Phil. Soc.* **147** (2009), pp 95–114.

[23] Koudou, A. E. and Ley, C. characterisations of GIG laws: a survey complemented with two new results. *Probab. Surv.* **11** (2014), pp. 161–176.

[24] Kumar, A. N. and Upadhye, N. S. On perturbations of Stein operator. *Commun. Stat. Theory* **46** (2017), pp. 9284–9302.

[25] Kusuoka, S. and Tudor, C. A. Stein's method for invariant measures of diffusions via Malliavin calculus. *Stoch. Proc. Appl.* **122** (2012), pp. 1627–1651.

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

[27] Luk, H. *Stein's Method for the Gamma Distribution and Related Statistical Applications.* PhD thesis, University of Southern California, 1994.

[28] Luke, Y. L. *The Special Functions and their Approximations, Vol. 1*, Academic Press, New York, 1969.

[29] McKeague, I., Peköz, E. and Swan, Y. Stein's method, many interacting worlds and quantum mechanics. To appear in *Bernoulli*, 2018+.

[30] Nourdin, I. and Peccati, G. Normal approximations with Malliavin calculus: from Stein's method to universality. Vol. 192. Cambridge University Press, 2012.

[31] Nourdin, I., Peccati, G. and Swan, Y. Entropy and the fourth moment phenomenons. *J. Funct. Anal.*, **266** (2014), pp. 3170–3207.

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

[33] Olver, F. W. J., Lozier, D. W., Boisvert, R. F. and Clark, C. W. *NIST Handbook of Mathematical Functions.* Cambridge University Press, 2010.

[34] Peköz, E., Röllin, A. and Ross, N. Degree asymptotics with rates for preferential attachment random graphs. *Ann. Appl. Probab.* **23** (2013), pp. 1188–1218.

[35] Peköz, E., Röllin, A. and Ross, N. Generalized gamma approximation with rates for urns, walks and trees. *Ann. Probab.* **44** (2016), pp. 1776–1816.

[36] Pike, J. and Ren, H. Stein's method and the Laplace distribution. *ALEA Lat. Am. J. Probab. Math. Stat.* **11** (2014), pp. 571-587.

[37] Ross, N. Fundamentals of Stein's method. *Probab. Surv.* **8** (2011), pp. 210-293.

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

[39] Springer, M. D. and Thompson, W. E. The distribution of independent random variables. *SIAM J. Appl. Math.* **14** (1966), pp. 511–526.

[40] Springer, M. D. and Thompson, W. E. The distribution of products of Beta, Gamma and Gaussian random variables. *SIAM J. Appl. Math.* **18** (1970), pp. 721–737.

[41] Stein, C. A bound for the error in the normal approximation to the the distribution of a sum of dependent random variables. In *Proc. Sixth Berkeley Symp. Math. Statis. Prob.* (1972), vol. 2, Univ. California Press, Berkeley, pp. 583–602.

[42] Stein, C. *Approximate Computation of Expectations.* IMS, Hayward, California, 1986.

[43] Xu, L. Approximation of stable law in Wasserstein-1 distance by Stein's method. arXiv:1709.00805, 2018.