2011/38

Univariate and Multivariate Chen-Stein Characterizations - a Parametric Approach

Christophe LEY
Yvik SWAN
Univariate and multivariate Chen-Stein characterizations – a parametric approach

Christophe Ley\textsuperscript{1} and Yvik Swan\textsuperscript{2}

Department of Mathematics
Université Libre de Bruxelles
Campus Plaine – CP210
1050 Brussels
Belgium

e-mail: chrisley@ulb.ac.be
yvswan@ulb.ac.be

Abstract: We provide a general framework for characterizing families of
(univariate, multivariate, discrete and continuous) distributions in terms
of a parameter of interest. We show how this allows for recovering known
Chen-Stein characterizations, and for constructing many more. Several ex-
amples are worked out in full, and different potential applications are dis-
cussed.

AMS 2000 subject classifications: Primary 20B10; secondary 62E17.
Keywords and phrases: Characterization theorems, Chen-Stein charac-
terizations, location and scale parameters, parameter of interest.

1. Introduction

In probability and statistics, a characterization theorem occurs whenever a given
law is the only one which satisfies a certain property. A famous example of such
a result is due to Carl Friedrich Gauss (see [15]), who proved that the normal is
the only continuous distribution for which the sample mean is always (that is,
for all samples) the maximum likelihood estimator of the location parameter.
Over the years, a large number of characterizations have been uncovered for
a wide range of (continuous, discrete, univariate, multivariate) distributions.
Although these theorems take on multiple guises, they all share a similar flavor
with Gauss’ result. They also pave the way for generalizations and extensions
which not only deepen our understanding of the laws under scrutiny, but also
allow to build bridges between different mathematical theories which would,
without them, seem unrelated.

Aside from their evident theoretical appeal, characterization theorems have
also proven to be of great use in statistics (see, for instance, [25], [23], [24] or

\textsuperscript{1}Supported by a Mandat d’Aspirant from the Fonds National de la Recherche Scientifique,
Communauté française de Belgique. Christophe Ley is also member of E.C.A.R.E.S.
\textsuperscript{2}Supported by a Mandat de Chargé de recherche from the Fonds National de la Recherche
Scientifique, Communauté française de Belgique.
C. Ley and Y. Swan. /Univariate and multivariate Chen-Stein characterizations

[26]), and in applied probability (see references given hereafter). Of particular importance in the latter field are the Chen-Stein characterizations, about which this article is concerned.

In a seminal paper, [31], Charles Stein used a characterization of the Gaussian distribution in terms of a certain operator to derive explicit rates of convergence in central limit theorems for dependent summands. Following up on this idea, Louis Chen constructed, in [8], a similar characterizing operator for the Poisson law which he used to derive convergence rates for Poisson approximation. Both papers are by now classical, and extensions of their methodology have resulted in what is now known as the Chen-Stein method (see [4] or [10] for a comprehensive overview).

The Chen-Stein method is a technique from probability theory for obtaining bounds on the distance, with respect to a given probability metric, between an unknown probability distribution (typically that of a sum of random variables) and a given probability distribution (the target distribution). The first (and easiest) step in this method consists in determining a suitable characterization of the target distribution. Chen and Stein provide such characterizations for the Poisson and the Gaussian laws. A groundbreaking step towards a generalization of their method for non-Gaussian or non-Poisson target distributions is due to Barbour [1] who showed that “in order to choose an appropriate [characterization] for less common limits, one possibility is to look for a Markov process [...] whose equilibrium distribution is the one required”. Barbour’s result was extended by Götze in [16]. The resulting construction, known as the generator approach (see, e.g., [22] or [29]), not only allows for characterizing a wide class of univariate distributions but can also be extended to the multivariate setting, and even to more abstract spaces. This is, however, not the only way to construct characterizations and Stein himself proposes a general density approach for constructing other types of characterizations (see [32], [11] and [33]).

Starting from a given characterization, the Chen-Stein method consists in a careful combination of many concepts such as, e.g., Stein equations, magic factors, exchangeable pairs, size-biased distributions, etc. (see [2] and [3] for an introduction). It has proven, over the last few decades, to be an extremely powerful tool for obtaining rates of convergence in stochastic approximation problems coming from a wide variety of fields. For the sake of illustration, we cite a number of recent successes which are relevant to this work. Götze and Tikhomirov use a characterization of the semi-circular law to obtain rates of convergence for spectra of random matrices with martingale structure (see [17]). Chatterjee, Fulman and Röllin use two different characterizations of the exponential distribution to obtain general results for convergence towards an exponential distribution (see [5]); they illustrate the applications of their methodology in a study of the spectrum of graphs with a non-normal limit. In [33], Stein, Diaconis, Holmes and Reinert obtain a characterization of a wide family of distributions which they use in the analysis of simulations. Eichelsbacher and Löwe [12] and Chatterjee and Shao [7] have recourse to the latter characterization in order to obtain general non-Gaussian approximation theorems, relevant for example in the field of statistical mechanics. Extensions to a multivariate setting are
also available for the multivariate Gaussian law (see, for instance, [20], [6], or [30]). There exists a uniform treatment of the univariate discrete case, by means of Gibbs measures, which can be found in [13]. Extensions to continuous time processes are currently the object of active research (see [28] or [27] and the references therein). This list is, of course, not exhaustive.

To the best of our understanding, there exists no general framework in which all the above characterizations could be seen as different instances of a unique phenomenon. The purpose of this work is to propose one such framework.

Our main result allows for recovering many of the characterizations used in the literature and equips the reader with a simple tool for constructing new ones. More importantly, it also provides a parametric interpretation of these results in terms of the parameters of the target distribution. For instance we show how Stein’s initial characterization and its extensions obtained via the density approach (see [33]), as well as those for exponential families (see [23] and [24]), are, in fact, location-based, i.e. they are characterizations one obtains when considering the target distribution to be of the form $g(x - \mu)$ for $g$ a density and $\mu \in \mathbb{R}$ a location parameter. We also provide scale-based characterizations and characterizations for discrete distributions. We furthermore provide a method for constructing characterizations of distributions in terms of any of their parameters. It appears that such parameter-based characterizations have never before been considered in the literature.

The outline of the paper is the following. In Section 2 we state and prove our main result, Theorem 2.1, which provides a general and simple characterization theorem for a very broad family of (discrete and continuous, univariate and multivariate) distributions. In Section 3 we illustrate the consequences of Theorem 2.1 by providing characterizations for important classes of parametric distributions, namely the location and the scale families, as well as a characterization for discrete distributions. We also apply our findings in a number of illustrative examples. In Section 4, we discuss some potential applications of our result, and provide several directions for future work. Finally, an Appendix collects the more technical proofs.

2. A general construction

In this section we present the main result of this paper, Theorem 2.1, which provides a unified framework for constructing Chen-Stein characterizations – by means of a characterizing class of test functions and a characterizing operator – for univariate, multivariate, discrete and continuous distributions. We show, hereby, that all these results allow for an interpretation in terms of a parameter of interest of the target distribution.

We first need to clearly identify the notations and vocabulary which will be used in this paper. Throughout, we let $k, p \in \mathbb{N}_0$ and consider the two measure spaces $(\mathcal{X}, \mathcal{B}_X, m_X)$ and $(\Theta, \mathcal{B}_\Theta, m_\Theta)$, where $\mathcal{X}$ is either $\mathbb{R}^k$ or $\mathbb{Z}^k$, where $\Theta$ is a subset of $\mathbb{R}^p$ whose interior is non-empty, where $m_X$ is either the Lebesgue measure or the counting measure, depending on the nature of $\mathcal{X}$, where $m_\Theta$ is
the Lebesgue measure, and where $B_X$ and $B_\Theta$ are the corresponding $\sigma$-algebras.

In this setup we disregard the case of discrete parameter spaces (as in, e.g., the discrete uniform); such distributions are shortly addressed in Remark 2.4 at the end of the current section.

Consider a couple $(X, \Theta)$ equipped with the corresponding $\sigma$-algebras and measures. We say that the measurable function $g : X \times \Theta \to \mathbb{R}^+$ forms a family of $\theta$-parametric densities, denoted by $g(\cdot; \theta)$, if $\int_{X} g(x; \theta) dm_X(x) = 1$ for all $\theta \in \Theta$. In this case we call $\theta$ the parameter of interest for $g$. When $X = \mathbb{R}^k$, corresponding to the absolutely continuous case, the mapping $x \mapsto g(x; \theta)$ is, for all $\theta \in \Theta$, a probability density function evaluated at the point $x \in \mathbb{R}^k$.

When $X = \mathbb{Z}^k$, corresponding to the discrete case, $g(x; \theta)$ is the probability mass associated with $x \in \mathbb{Z}^k$ and $g(\cdot; \theta)$ therefore maps $\mathbb{Z}^k$ onto $[0, 1]$. This unified terminology will allow us to treat absolutely continuous and discrete distributions in one common framework. For the sake of simplicity, we rule out mixed distributions.

**Example 2.1.** $\theta$-parametric densities are ubiquitous in probability and statistics. Taking $g : \mathbb{Z} \times \mathbb{R}^+ \to [0, 1] : (x, \lambda) \mapsto e^{-\lambda x}/x! \, \mathbb{I}_A(x)$, where $\mathbb{I}_A(\cdot)$ stands for the indicator function of some set $A \in B_X$, we see the density of a Poisson $\mathcal{P}(\lambda)$ distribution as a $\lambda$-parametric density. Taking $g : \mathbb{R} \times (\mathbb{R} \times \mathbb{R}^+ ) \to \mathbb{R}^+ : (x, (\mu, \sigma)^') \mapsto (2\pi \sigma^2)^{-1/2} e^{-(x-\mu)^2/(2\sigma^2)}$, we see the density of a Gaussian $\mathcal{N}(\mu, \sigma)$ distribution as a $(\mu, \sigma)$-parametric density. If, in the Gaussian case, the scale is known (and set to $\sigma_0$), one is then only interested in the location parameter $\mu$. Taking $\tilde{g} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+ : (x, \mu) \mapsto \tilde{g}(x; \mu) = g(x; (\mu, \sigma_0)^')$, we see the density of a Gaussian $\mathcal{N}(\mu, \sigma_0)$ distribution as a $\mu$-parametric density. Likewise, one can see the density of a uniform $U[a, b]$ distribution as an $(a, b)$-parametric density, an $a$-parametric density or a $b$-parametric density. In general, there are infinitely many ways to write the density of any given probability distribution as a $\theta$-parametric density for any given $\theta$. See for instance, on this issue, the discussion on the so-called natural parameters of the exponential family in [26].

Fix a couple $(X, \Theta)$ as above, endowed with their respective $\sigma$-algebras and measures. Throughout this paper, the densities we shall work with all belong to the class $\mathcal{G} := \mathcal{G}(X, \Theta)$ of $\theta$-parametric densities for which the mapping $\theta \mapsto g(\cdot; \theta)$ is differentiable in the sense of distributions. Such distributions may have a bounded support possibly depending on the parameter $\theta$; we will denote this support by $S_\theta := S_\theta(g)$, be it dependent on $\theta$ or not.

With this in hand, we are ready to define the two fundamental concepts of this paper. These (a class of functions and an operator) mirror notions already present in the literature on Chen-Stein characterizations.

**Definition 2.1.** Let $\theta_0$ be an interior point of $\Theta$ and let $g \in \mathcal{G}$. We define the class $\mathcal{F}(g; \theta_0)$ as the collection of test functions $f : X \times \Theta \to \mathbb{R}$ such that the following three conditions are satisfied in some neighborhood $\Theta_0 \subset \Theta$ of $\theta_0$.

Condition (i) : there exists $c_f \in \mathbb{R}$ such that $\int_X f(x; \theta) g(x; \theta) dm_X(x) = c_f$ for all $\theta \in \Theta_0$. 
Condition (ii) : the mapping $\theta \mapsto f(\cdot; \theta)g(\cdot; \theta)$ is differentiable in the sense of distributions over $\Theta_0$.

Condition (iii) : there exist $p$ $m_X$-integrable functions $h_i : X \to \mathbb{R}^+$, $i = 1, \ldots, p$, such that $|\partial_\theta_i (f(x; \theta)g(x; \theta))| \leq h_i(x)$ over $X$ for all $i = 1, \ldots, p$ and for all $\theta \in \Theta_0$.

Definition 2.2. Let $\theta_0$ be an interior point of $\Theta$. Also let $g$ and $F(g; \theta_0)$ be as above. We define the Chen-Stein operator $T_{\theta_0} := T_{\theta_0}(:, g) : F(g; \theta_0) \to X^*$ as

$$T_{\theta_0}(f, g)(x) = \frac{\nabla_\theta (f(x; \theta)g(x; \theta))|_{\theta = \theta_0}}{g(x; \theta_0)}.$$  

(2.1)

The operator defined by (2.1) requires some comments. If the support of $g(\cdot; \theta)$ is $X$ itself, then the operator is obviously well-defined everywhere. If, on the contrary, the density $g(\cdot; \theta)$ has support $S_0 \subset X$, then there is an ambiguity which we need to avoid. To this end we adopt the convention that, whenever an expression involves the division by an indicator function $\mathbb{1}_A$ for some $A \in \mathcal{B}_X$, we are, in fact, multiplying the expression by the said indicator function.

With this convention, writing out the operator in full (whenever the gradient $\nabla_\theta (f(x; \theta))|_{\theta = \theta_0}$ is well-defined on $X$) reads

$$T_{\theta_0}(f, g)(x) = \left( \nabla_\theta (f(x; \theta))|_{\theta = \theta_0} + f(x; \theta) \frac{\nabla_\theta (g(x; \theta))|_{\theta = \theta_0}}{g(x; \theta_0)} \right) \mathbb{1}_{S_{\theta_0}}(x).$$

Our convention not only guarantees that the Chen-Stein operator is well-defined but also that, for any test function $f$,-\$ part of this form, it becomes $T_{\theta_0}(f, g)(x) = -f_0(x) + xf_0(x).$

Example 2.2. (i) Let $X = \mathbb{R}$, $\Theta = \mathbb{R}$ and $g(x; \mu) = (2\pi)^{-1/2}e^{-(x-\mu)^2/2}$, the density of a univariate normal $N(\mu, 1)$ distribution. Clearly, $g$ belongs to $\mathcal{G}$ for all $\mu \in \mathbb{R}$ and its support $S_0 = \mathbb{R}$ is independent of $\mu$. Fix $\mu_0 = 0$ and consider functions of the form $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R} : (x, \mu) \mapsto f(x; \mu) := f_0(x - \mu)$, where $f_0 : \mathbb{R} \to \mathbb{R}$ is chosen such that $f \in F(g; 0)$. Restricting the operator $T_{\theta_0}$ to the collection of $f$’s of this form, it becomes

$$T_{\theta_0}(f, g)(x) = -f_0(x) + xf_0(x).$$

(ii) Let $X = \mathbb{Z}$, $\Theta = \mathbb{R}_0^+$ and $g(x; \lambda) = e^{-\lambda}x^2/f! \mathbb{1}_{\mathbb{N}}(x)$, the density of a Poisson $\mathcal{P}(\lambda)$ distribution. Clearly, $g$ belongs to $\mathcal{G}$ for all $\lambda \in \mathbb{R}_0^+$ and its support $S_\lambda = \mathbb{N}$ is independent of $\lambda$. Fix $\lambda = \lambda_0$ and consider functions of the form $f : \mathbb{Z} \times \mathbb{R}_0^+ \to \mathbb{R} : (x, \lambda) \mapsto f(x; \lambda) := e^{\lambda}f_0(x+1)/(x+1) - f_0(x)$, where $f_0 : \mathbb{Z} \to \mathbb{R}$ is chosen such that $f \in F(g; \lambda_0)$. Restricting the operator $T_{\theta_0}$ to the collection of $f$’s of this form, it becomes

$$T_{\theta_0}(f, g)(x) = e^{\lambda_0} \left( f_0(x+1) - \frac{x}{\lambda_0}f_0(x) \right) \mathbb{1}_{\mathbb{N}}(x).$$
Among densities \( g \in G \), those which satisfy the following (local) regularity assumption at a given interior point \( \theta_0 \in \Theta \) will play a particular role.

Assumption A: there exists a rectangular bounded neighborhood \( \Theta_0 \subset \Theta \) of \( \theta_0 \) and a \( m_X \)-integrable function \( h : X \to \mathbb{R}^+ \) such that \( g(x; \theta) \leq h(x) \) over \( X \) for all \( \theta \in \Theta_0 \).

This assumption is weak, and is satisfied for example as soon as the target density is bounded over its support. It does, nevertheless, exclude some well-known distributions such as, e.g., the arcsine distribution.

With these notations, we are ready to state and prove our general characterization theorem.

**Theorem 2.1.** Let \( g \in G \), let \( Z_\theta \) be distributed according to \( g(\cdot; \theta) \), and let \( X \) be a random vector taking values on \( X \). Fix an interior point \( \theta_0 \in \Theta \). Then the following two assertions hold.

1. If \( X \overset{\mathcal{L}}{=} Z_{\theta_0} \), then \( \mathbb{E}[T_{\theta_0}(f, g)(X)] = 0 \) for all \( f \in F(g; \theta_0) \).

2. If \( g \) also satisfies Assumption A at \( \theta_0 \) and if \( \mathbb{E}[T_{\theta_0}(f, g)(X)] = 0 \) for all \( f \in F(g; \theta_0) \), then \( X \mid X \in S_{\theta_0} \overset{\mathcal{L}}{=} Z_{\theta_0} \).

The first statement in Theorem 2.1 is standard; it implies that in order to obtain a Chen-Stein operator for a given \( \theta \)-parametric density \( g \) at a point \( \theta_0 \), it suffices to find a collection of functions \( f \) such that the conditions in Definition 2.1 hold and then apply the operator given in Definition 2.2. As we will show in the next section, this allows for recovering many well-known Chen-Stein operators, and for constructing many more. The second statement is also quite standard whenever \( S_{\theta_0} = X \). If \( S_{\theta_0} \subset X \), then things are slightly more tricky. Indeed, in this case, equation (2.2) does not imply that the law of \( X \) is necessarily that of \( Z_{\theta_0} \), but rather that if the distribution of \( X \) has support \( S_{\theta_0} \) and if \( X \) satisfies \( \mathbb{E}[T_{\theta_0}(f, g)(X)] = 0 \) (on \( S_{\theta_0} \) by definition of \( T_{\theta_0}(f, g) \)) for all \( f \in F(g; \theta_0) \), then \( X \) is distributed according to \( g(\cdot; \theta_0) \). This is in accordance with all other results of this form.

**Proof.** (1) Since Condition (iii) allows for differentiating w.r.t. \( \theta \) under the integral in Condition (i) and since differentiating w.r.t. \( \theta \) is allowed thanks to Condition (ii), the claim follows immediately.

(2) First suppose that \( p = 1 \), and fix \( \Theta_0 \subset \Theta \), a bounded (rectangular) neighborhood of \( \theta_0 \) on which \( g \) satisfies Assumption A at \( \theta_0 \). Define, for \( A \in \mathcal{B}_X \), the mapping

\[
f_A : X \times \Theta_0 \to \mathbb{R} : (x, \theta) \mapsto \frac{1}{g(x; \theta)} \int_{\theta_0}^{\theta} l_A(x; u, \theta) g(x; u) dm_\Theta(u) \tag{2.3}
\]

with

\[
l_A(x; u, \theta) := (1_A(x) - P(Z_u \in A \mid Z_u \in S_{\theta})) 1_{S_{\theta}}(x),
\]

C. Ley and Y. Swan./Univariate and multivariate Chen-Stein characterizations 6

Among densities \( g \in G \), those which satisfy the following (local) regularity assumption at a given interior point \( \theta_0 \in \Theta \) will play a particular role.
where

$$P(Z_u \in B) = \int_X I_B(x)g(x; u)dm_X(x)$$

for $B \in B_X$. Note that, for the event $[Z_u \in S_\theta]$ to have a non-zero probability, it is crucial to work in a neighborhood $\Theta_0$ rather than in $\Theta$; clearly, this event is always true when $S_\theta$ does not depend on $\theta$. To see that $f_A$ belongs to $F(g; \theta_0)$, first note that

$$\int_X f_A(x; \theta)g(x; \theta)dm_X(x) = \int_X \int_{\theta_0}^\theta l_A(x; u, \theta)g(x; u)dm_\Theta(u)dm_X(x)$$

$$= \int_{\theta_0}^\theta \int_X l_A(x; u, \theta)g(x; u)dm_X(x)dm_\Theta(u),$$

where the last equality follows from Fubini’s theorem, which can be applied for all $\theta \in \Theta_0$, since in this case there exists a constant $M$ such that

$$\int_{\theta_0}^\theta \int_X |l_A(x; u, \theta)|g(x; u)dm_X(x)dm_\Theta(u) \leq 2|\theta - \theta_0| \leq M$$

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

$$\int_X l_A(x; u, \theta)g(x; u)dm_X(x)$$

$$= P(Z_u \in A \cap S_\theta) - P(Z_u \in A \mid Z_u \in S_\theta)P(Z_u \in S_\theta)$$

$$= 0.$$

Hence $f_A$ satisfies Condition (i). Condition (ii) is easily checked. Regarding Condition (iii), one sees that

$$\partial_t (f_A(x; t)g(x; t))|_{t=\theta} = l_A(x; \theta, \theta)g(x; \theta) + H(x; \theta),$$

(2.4)

with $H(x; \theta)$ a function whose complete expression is provided in the Appendix. As shown there, it is easy to bound $H(x; \theta)$ uniformly in $\theta$ over $\Theta_0$ by a $m_X$-integrable function. Moreover, Assumption A guarantees that the same holds for $l_A(x; \theta, \theta)g(x; \theta)$. Hence $f_A$ satisfies Condition (iii). Wrapping up, we have thus proved that $f_A \in F(g; \theta_0)$. The conclusion follows, since $H(x; \theta_0) = 0$ for all $x \in X$ (see the Appendix) and since, by hypothesis,

$$E[T_{\theta_0}(f_A, g)(X)] = E[I_{A \cap S_{\theta_0}}(X) - P(Z_{\theta_0} \in A)I_{S_{\theta_0}}(X)] = 0.$$

Next suppose that $p > 1$. Let $\theta_0 := (\theta_0^1, \ldots, \theta_0^p)$ and fix $\Theta_0 := \Theta_0^1 \times \cdots \times \Theta_0^p$ a bounded (rectangular) neighborhood of $\theta_0$ on which $g$ satisfies Assumption A at $\theta_0$. Define, for all $j = 1, \ldots, p$ and for all $A \in B_X$, the mappings

$$\bar{\theta}_j^i : \Theta_0 \rightarrow \Theta_0 : u \mapsto (\theta_0^1, \ldots, \theta_0^{i-1}, u, \theta_0^{i+1}, \ldots, \theta_0^p)$$

and

$$f_\lambda^j : X \times \Theta_0 \rightarrow \mathbb{R} : (x, \theta) \mapsto \frac{1}{g(x; \theta)} \int_{\bar{\theta}_j^0}^{\bar{\theta}_j^i} l_A(x; u, \theta)g(x; \bar{\theta}_j^i(u))dm_\Theta(u),$$
with
\[ f_A^p(x; u, \theta^j) := \left( I_A(x) - P \left( Z^j_u \in A \mid Z^j_u \in S_{\theta^j(u)} \right) \right) \| S_{\theta^j(u)}(x), \]

where
\[ P(Z^j_u \in B) := \int_X I_B(x) g(x; \bar{\theta}^j(u)) dm_X(x) \]

for \( B \in B_X \). The \( p \)-variate equivalent of the function \( f_A \) in (2.3) is given by
\[ f_A^p(x; \theta) := \sum_{j=1}^p f_A^j(x; \theta). \]

Along the same lines as for the special case \( p = 1 \), Conditions (i)-(iii) are now easily seen to be satisfied by \( f_A^p \) (we draw the reader’s attention to the fact that the rectangular nature of the neighborhood \( \Theta_0 \) is important in order to ensure Condition (iii)). The result readily follows.

**Remark 2.1.** Nowhere in the proof did we need to specify whether the random vector \( X \) is univariate (for \( k = 1 \)) or multivariate (for \( k > 1 \)).

**Remark 2.2.** When \( p > 1 \), the (vectorial) operator \( T_{\theta_0}(f, g) \) contains, in a sense, \( p \) different characterizations of the \( \theta = (\theta^1, \ldots, \theta^p) \)-parametric density \( g \) at \( \theta_0 \). The requirements (in this formulation of the result) on the test functions \( f \) are, perhaps, unnecessarily stringent. Indeed, setting \( \theta^{(q)} := (\theta^{i_1}, \ldots, \theta^{i_q}) \) for \( 1 \leq i_1 \leq \ldots \leq i_q \leq p \), we can obviously consider \( g \) as a \( \theta^{(q)} \)-parametric density. The corresponding \( q \)-dimensional sub-vector of \( T_{\theta_0}(f, g) \) also gives rise to a (vectorial) Chen-Stein operator for which the conclusions of Theorem 2.1 also hold at \( \theta_0 \), this time with a possibly larger class of test functions \( f \) (thanks to the weakening of the requirements imposed by Condition (iii)). In particular, taking \( q = 1 \), we obtain \( p \) distinct one-dimensional characterizations of \( g \) at \( \theta_0 \). This might be very helpful in approximation theorems concerning \( g \); see Section 4 for more details.

**Remark 2.3.** Note that both implications in Theorem 2.1 are obtained at fixed \( \theta_0 \in \Theta \). We attract the reader’s attention to the fact that all our calculations and manipulations, as well as all the conditions on the functions at play, are consequently local around \( \theta_0 \).

**Remark 2.4.** All the definitions and arguments above can be extended to encompass distributions with a discrete parameter space \( \Theta \) (such as, e.g., the discrete uniform). For this it suffices, in a sense, to replace the derivatives and integrals by forward (or backward) differences and summations, respectively. Although it is easy to obtain Chen-Stein operators by this means, determining the exact conditions under which the theorem holds nevertheless requires some care, since in this case there arise problems which originate in the interplay between the support of the target density and the parameter of interest. Because of these (structural) intricacies, working out explicit conditions on the target density in this framework appears to be a rather sterile exercise, which is perhaps better suited to ad hoc case by case arguments. This issue will no longer be addressed within the present paper.
The first statement of Theorem 2.1 can be seen as a user-friendly Chen-Stein operator-producing mechanism, since any subclass \( \tilde{F}(g; \theta_0) \subset F(g; \theta_0) \) yields a left-right implication, i.e. an implication of the form

\[
X \sim g(\cdot; \theta_0) \implies E[\mathcal{T}_{\theta_0}(f, g)(X)] = 0 \text{ for all } f \in \tilde{F}(g; \theta_0).
\]

This raises some important questions. Indeed, consider for instance the two operators provided in Example 2.2. As it turns out, both these operators have proven to be extremely useful in applications and their properties are fundamental in the history of the Chen-Stein method. However, as already noted by a number of authors before us, they are by no means the only such operators for the Gaussian or the Poisson distribution; in our framework they are just two particular instances of equation (2.1) restricted to certain very specific forms of test functions. A natural question is therefore that of whether there exist other subclasses of test functions for which the corresponding operators would also be useful in applications. It is possible that this question does not allow for a fully satisfactory answer. More precisely it is possible that, for any given problem, there is no \textit{a priori} reason why a given operator would yield better rates of convergence than any other, and perhaps in each problem a careful combination of different characterizations (\textit{à la} Chatterjee, Fulman and Röllin [5]) would be fruitful and would allow for obtaining better results than those obtained by focusing on a single characterization alone. We will return to this issue in Section 4.

In any case it seems intuitively clear that, in order for a subclass and the corresponding operator to be of practical use, they need to characterize the law under consideration, that is, we should have the relationship

\[
X \sim g(\cdot; \theta_0) \iff E[\mathcal{T}_{\theta_0}(f, g)(X)] = 0 \text{ for all } f \in \tilde{F}(g; \theta_0),
\]

where the right-left implication is to be understood in the sense of (2.2) in case \( S_{\theta_0} \) is a strict subset of \( X \). Constructing such subclasses, which we call \( \theta \)-\textit{characterizing} for \( g \) at \( \theta_0 \), is relatively easy. Indeed it suffices to adjoin the function \( f_A \) defined in (2.3) to any collection (even empty) of test functions which satisfy the three conditions in Definition 2.1. Such an approach is, however, of limited interest and, moreover, does not allow for clearly identifying the form of the corresponding operators. We therefore suggest a more constructive approach, which we describe in detail in the next section.

3. Applications

In this section we provide a general “recipe” which allows for constructing \( \theta \)-characterizing subclasses with well-identified operators. We apply our method to build general characterizations for location families, scale families and discrete distributions. Many well-known Chen-Stein characterizations fall under the umbrella of these results. We also show how our method can be applied to obtain more unusual characterizations.
3.1. A recipe

For the sake of simplicity, we let \( k = p = 1 \). Fix \( \theta_0 \in \Theta \) and choose \( g \in \mathcal{G} \) which satisfies Assumption A at \( \theta_0 \). In order to construct a \( \theta \)-characterizing subclass \( \mathcal{F}(g; \theta_0) \subset \mathcal{F}(g; \theta_0) \), we suggest the following method.

**Step 1:** Consider Condition (i) in Definition 2.1, which requires that we have
\[
\int_X f(x; \theta) g(x; \theta) dm_X(x) = c_f
\]
for \( c_f \in \mathbb{R} \). In many cases, the interaction between the variable \( x \) and the parameter \( \theta \) within the density \( g \) allows to determine a favored family of test functions \( \tilde{f}_0(x; \theta) \) which satisfy this condition. Moreover, these functions are usually expressible as \( \tilde{f}_0(x; \theta) = \tilde{T}(f_0; \theta)(x) \), with \( f_0 \in \mathcal{X}^* \) and \( \tilde{T} : \mathcal{X}^* \times \Theta \rightarrow (\mathcal{X} \times \Theta)^* \).

**Step 2:** For \( \tilde{T} \) and \( f_0 \) as given in Step 1, define the **exchanging operator** \( T : \mathcal{X}^* \times \Theta \rightarrow (\mathcal{X} \times \Theta)^* \) as a transformation which satisfies the **exchangeability condition**
\[
\partial_{\theta} \left( \tilde{T}(f_0; \theta)(x) g(x; \theta) \right) \bigg|_{\theta = \theta_0} = \partial_{y} (T(f_0; \theta_0)(y) g(y; \theta_0)) \bigg|_{y = x} \quad (3.1)
\]
over \( \mathcal{X} \), where \( \partial_{\theta} \) either means the derivative in the sense of distributions or the discrete (forward or backward) difference, and we hereby implicitly require that \( T \) is such that the derivative on the rhs of (3.1) is well-defined over \( \mathcal{X} \).

**Step 3:** Define the class \( \mathcal{F}_0 := \mathcal{F}_0(g; \theta_0) \) as the collection of all functions \( f_0 \in \mathcal{X}^* \) such that \( \tilde{T}(f_0; \theta) \in \mathcal{F}(g; \theta_0) \). Note that we therefore have the (new) left-right implication
\[
X \sim g(·; \theta_0) \implies E \left[ \frac{\partial_{y} (T(f_0; \theta_0)(y) g(y; \theta_0))}{g(X; \theta_0)} \bigg|_{y = X} \right] = 0 \text{ for all } f_0 \in \mathcal{F}_0.
\]

**Step 4:** Solve the **Chen-Stein equation**
\[
\partial_{\theta} \left( T(f_0^A; \theta_0)(y) g(y; \theta_0) \right) \bigg|_{y = x} = l_A(x; \theta_0, \theta_0) g(x; \theta_0) \quad (3.2)
\]
where \( l_A(x; \theta_0, \theta_0) \) is as in the proof of Theorem 2.1. If \( T(·; \theta_0) \) is invertible, it then suffices to check whether the corresponding \( f_0^A \) belongs to \( \mathcal{F}_0 \) in order to obtain the characterization
\[
X \sim g(·; \theta_0) \iff E \left[ \frac{\partial_{y} (T(f_0; \theta_0)(y) g(y; \theta_0))}{g(X; \theta_0)} \bigg|_{y = X} \right] = 0 \text{ for all } f_0 \in \mathcal{F}_0,
\]
where the right-left implication is to be understood, as before, in the sense of (2.2) in case the support \( S_{\theta_0} \) of \( g(·; \theta_0) \) is a strict subset of \( \mathcal{X} \).
The resulting $\theta$-characterizing subclass $\tilde{F}(g; \theta_0)$ is none other than the collection $\{\tilde{T}(f_0; \theta) \mid f_0 \in F_0\} \cup \{f_A\}$; this collection not only has the desired properties, but also is accompanied with a well-identified Chen-Stein operator. In the sequel it will be more convenient to state our results in terms of $F_0$ rather than in terms of $\tilde{F}(g; \theta_0)$. This is in accordance with all other results of this form.

There are a number of ways in which one can extend the method presented above to the cases $k > 1$ and $p > 1$. Also, for given $\theta_0$ and $\theta$-parametric density $g$, the choice of class $F_0$ and exchanging operator $T$ is not unique. Moreover, determining straightforward minimal conditions on the $f_0$ for the characterization to hold seems to be impossible without making further regularity assumptions on the target density $g$. These considerations entail that it is perhaps more fruitful to tackle different $\theta$-parametric densities with ad hoc arguments. There are nevertheless important instances in which one can obtain general results with relative ease. To this end consider the following assumption on univariate $\theta$-parametric densities.

Assumption B : there exists $x_0 \in \mathcal{X}$ such that

$$\left|\int_{\mathcal{X}} \left(\int_{x_0}^{x} l_A(y; \theta_0, \theta_0)g(y; \theta_0)dm_X(y)\right)1_{S_{\theta_0}}(x)dm_X(x)\right| < \infty$$

for all $A \in \mathcal{B}_X$, where $l_A(y; \theta_0, \theta_0)$ is defined as in the proof of Theorem 2.1.

This is a condition on the tails of the density $g(\cdot; \theta_0)$ which is, for instance, satisfied by the Gaussian and the exponential distributions (while the latter is evident, see for the former [9] page 4). As we will see, Assumption B is useful for determining general characterization results in location and scale models.

### 3.2. Location-based characterizations

In this subsection we apply the method described in Section 3.1 to study laws whose parameter of interest is a location parameter.

**Corollary 3.1.** Let $k = p = 1$ and $\mathcal{X} = \mathbb{R} = \Theta$, and fix $\mu_0 \in \Theta$. Define $\mathcal{G}_{loc}$ as the collection of densities $g_0 : \mathcal{X} \to \mathbb{R}^+$ with support $S \subset \mathcal{X}$ such that the $\mu$-parametric density $g(x; \mu) = g_0(x - \mu)$ belongs to $\mathcal{G}$ and satisfies Assumptions A and B at $\mu_0$. Let $\Theta_0 \subset \Theta$ be as in Assumption A, and define $F_0 := F_0(g_0; \mu_0)$ as the collection of all $f_0 : \mathcal{X} \to \mathbb{R}$ such that

- Condition ($\mu$-i) : $\left|\int_{\mathcal{X}} f_0(x)g_0(x)dm_X(x)\right| < \infty$,
- Condition ($\mu$-ii) : the mapping $x \mapsto f_0(x)g_0(x)$ is differentiable in the sense of distributions over $\mathcal{X}$
- Condition ($\mu$-iii) : there exists a $m_X$-integrable function $h : \mathcal{X} \to \mathbb{R}^+$ such that $\left|\partial_y(f_0(y - \mu)g_0(y - \mu))\right|_{y=x} \leq h(x)$ over $\mathcal{X}$ for all $\mu \in \Theta_0$. 


Then $\mathcal{F}_0$ is $\mu$-characterizing for $g_0$ at $\mu_0$, with $\mu$-characterizing operator

$$
\mathcal{T}_{\mu_0}(f_0, g_0) : \mathcal{X} \to \mathcal{X} : x \mapsto -\frac{\partial_y (f_0(y - \mu_0)g_0(y - \mu_0))}{g_0(x - \mu_0)}|_{y = x}.
$$

(3.3)

A proof, which is a direct application of the method described in Section 3.1, is provided in the Appendix.

The operator in (3.3) – as well as the conditions on the densities and the conditions on the test functions $f_0$ – differ slightly from those already available in the literature. These differences require some comments. For example, in [33], a similar characterization is given for random variables whose density $g$ is positive, regular on an interval $[a, b]$, allows a (strong) derivative $\dot{g}$ which is also regular on $(a, b)$ and such that $\dot{g}/g$ is also regular on $(a, b)$ (by regular we mean all functions which are bounded and have at most countably many discontinuity points on their support). Their result is given in terms of the characterizing operator $\dot{f}(x) + \frac{\dot{g}(x)}{g(x)} f(x) + f(a^+)g(a^+) - f(b^-)g(b^-)$. The discrepancy between this operator and ours is mainly due to a different interpretation of derivatives. For random variables with support $S$ which is a strict subset of $\mathbb{R}$, we take derivatives (in the sense of distributions) on $\mathbb{R}$ of a function of the form $g(x)\mathbb{1}_S(x)$ whereas they take strong derivatives of $g$ on the interior of $S$ (and hence omit the factor $\mathbb{1}_S$ in their formulation). As is easily seen, the constant term $f(a^+)g(a^+) - f(b^-)g(b^-)$ originates in the derivative of the indicator function $\mathbb{1}_{[a,b]}(x)$.

Setting these minor differences aside, one readily sees that Corollary 3.1 contains a number of well-known univariate characterizations covered in the literature. For instance, taking $g(\cdot; \mu)$ to be the density of a $\mathcal{N}(\mu, 1)$ (which satisfies Assumptions A and B at $\mu_0 = 0$) we can use the operator provided in Example 2.2; Corollary 3.1 then leads to the famous Stein characterization of the standard normal distribution. Likewise, introducing an artificial location parameter $\mu$ within the exponential density with scale parameter $1$ (which, again, satisfies Assumptions A and B at $\mu_0 = 0$) leads to Chatterjee, Fulman and Röllin’s first characterization of the exponential distribution (see [5]). More generally, when $g$ belongs to the (continuous) exponential family, one easily sees how the same manipulations allow to retrieve the known characterizations (see, e.g., [23], [24] or [26]). See also [33] and [11] for more location-based characterizations.

Now consider the semi-circular law whose density is given by

$$
g_0(x - \mu) = \frac{2}{\pi\sigma^2} \sqrt{\sigma^2 - (x - \mu)^2} \mathbb{1}_{[-\sigma, \sigma]}(x - \mu),
$$

with $\mu \in \mathbb{R}$ being a location and $\sigma \in \mathbb{R}^+_0$ a known scale parameter. Note that, although we are in a location model with target density satisfying Assumptions A and B at all points $\mu_0 \in \mathbb{R}$, the derivative $g_0'(x - \mu)$ is not bounded at the edges of the support. Conditions (iii-ii) and (iii-iii) therefore entail some stringent requirements on the admissible class of test functions. In order to be able to read these requirements more easily, one way to proceed is to consider only $f_0$’s of the form $f_0(x) = f_1(x)(\sigma^2 - x^2)^r$, with $r > 1/2$. Sufficient conditions on $f_1$ for $f_0$ to belong to $\mathcal{F}_0$ are easy to provide (see [17] in the case $r = 1$ and $\sigma = 2$).
Finally consider a random $k$-vector $Z_{\mu_0}$ with $\mu_0 \in \mathbb{R}^k$ and density of the form $g(x; \mu) := g_0(x - \mu) = g_0(x_1 - \mu_1, x_2 - \mu_2, \ldots, x_k - \mu_k)$. Suppose, for the sake of simplicity, that the support of $g_0(x - \mu)$ does not depend on $\mu$ (i.e., $S = \mathcal{X} = \mathbb{R}^k$). One way to characterize such distributions at $\mu_0$ is to define, for fixed $x_2, \ldots, x_k$, the univariate $\mu^1$-parametric density $g_1(x_1; \mu_1) = g_0(x_1 - \mu_1^1, x_2 - \mu_0^2, \ldots, x_k - \mu_0^k)$. Requiring that $g_1 \in \mathcal{G}$ and satisfies Assumptions A and B at $\mu_0$, we easily determine a class of functions $\mathcal{F}_0^1$ as in Corollary 3.1 to obtain

$$X \overset{\mathcal{D}}{\sim} Z_{\mu_0} \iff \mathbb{E} \left[ \frac{\partial_y \left( f_0(y - \mu_0, X) g_0(y - \mu_0, X) \right) y = x_1}{g_0(X - \mu_0)} \right] = 0 \text{ for all } f_0 \in \mathcal{F}_0^1, \quad (3.4)$$

where we use the abuse of notations $(y - \mu_0, X) = (y - \mu_0^1, X_2 - \mu_0^2, \ldots, X_k - \mu_0^k)$ and $X - \mu_0 = (X_1 - \mu_0^1, X_2 - \mu_0^2, \ldots, X_k - \mu_0^k)$. The choice of $\mu^1$ as parameter of interest was of course for convenience only, and similar relationships hold for derivatives with respect to $x_2, \ldots, x_k$ as well. Moreover, when $Z_{\theta_0}$ has support $\mathcal{X}$ and independent marginals, one easily sees how to aggregate these different results and write out a class of functions $\mathcal{F}_0^{(k)}$ as in Corollary 3.1 to get

$$X \sim g(\cdot; \theta_0) \iff \mathbb{E} \left[ \frac{\nabla_y \left( f_0(y - \mu_0) g_0(y - \mu_0) \right) y = X}{g_0(X - \mu_0)} \right] = 0 \text{ for all } f_0 \in \mathcal{F}_0^{(k)} \quad (3.5)$$

We conclude this section by showing how (3.4) and (3.5) read in the Gaussian case. Here, setting $\mu_0 = 0 \in \mathbb{R}^k$ and plugging the multivariate Gaussian density $g(x; \mu, \Sigma)$ with $\Sigma$ a known symmetric positive definite $k \times k$ matrix into (3.4) we get, for $j = 1, \ldots, k$,

$$X \sim \mathcal{N}(0, \Sigma) \iff \mathbb{E} \left[ \nabla_{y_j} \left( f_0(y_j, X) \right) y_j = X_j - \sigma_j f_0(X) \right] = 0 \text{ for all } f_0 \in \mathcal{F}_0^j \quad (3.6)$$

where we use the notations $(y_j, X) = (X_1, \ldots, X_{j-1}, y_j, X_{j+1}, \ldots, X_k)$ and $\sigma_j := (\Sigma^{-1} X_j)^j = \sum_{i=1}^k (\Sigma^{-1})_{ij} X_i$. Moreover, when $\Sigma$ is the identity matrix $I_k$ we can use (3.5) to obtain

$$X \sim \mathcal{N}(0, I_k) \iff \mathbb{E} \left[ \nabla_y \left( f_0(y) \right) y = X f_0(X) \right] = 0 \text{ for all } f_0 \in \mathcal{F}_0^{(k)} \quad (3.7)$$

These characterizations of the multivariate Gaussian are, to the best of our knowledge, new. They are to be compared with existing results given, e.g., in [6] and [30], and in light of the discussion provided in Section 4.

### 3.3. Scale-based characterizations

In this subsection we apply the method described in Section 3.1 to study laws whose parameter of interest is a scale parameter.
Corollary 3.2. Let $k = p = 1$, $X = \mathbb{R}$ and $\Theta = \mathbb{R}_0^+$. Define $\mathcal{G}_{\text{sub}}$ as the collection of densities $g_0 : X \to \mathbb{R}^+$ with support $S \subset X$ such that the $\sigma$-parametric density $g(x; \sigma) = \sigma g_0(\sigma x)$ belongs to $\mathcal{G}$ and satisfies Assumptions $A$ and $B$ at $\sigma_0$. Let $\Theta_0 \subset \Theta$ be as in Assumption $A$, and define $\mathcal{F}_0 := \mathcal{F}_0(g_0; \sigma_0)$ as the collection of all $f_0 : X \to \mathbb{R}$ such that

$$\left| \int_X f_0(x)g_0(x)dm(x) \right| < \infty.$$  

Condition (σ-i) : $\left| \int_X f_0(x)g_0(x)dm(x) \right| < \infty$.

Condition (σ-ii) : the mapping $x \mapsto x f_0(x)g_0(x)$ is differentiable in the sense of distributions over $X$.

Condition (σ-iii) : there exists a $m_X$-integrable function $h : X \to \mathbb{R}^+$ such that $\left| \partial_y (y f_0(\sigma y)g_0(\sigma y)) \right|_{y=x} \leq h(x)$ over $X$ for all $\sigma \in \Theta_0$.

Then $\mathcal{F}_0$ is $\sigma$-characterizing for $g_0$ at $\sigma_0$, with $\sigma$-characterizing operator

$$\mathcal{T}_{\sigma_0}(f_0, g_0) : X \to X : x \mapsto \frac{\partial_y (y f_0(\sigma_0 y)g_0(\sigma_0 y))}{\sigma_0 g_0(\sigma_0 x)} \quad (3.8)$$

The proof of Corollary 3.2 is similar to that of Corollary 3.1, and hence is omitted.

As in the location case, this result can be extended in a number of ways to the multivariate setting. In the univariate setup, if $g$ is the exponential density with scale parameter $\lambda$ and if $\lambda_0$ is set to 1, we retrieve the second characterization of the $Exp(1)$ distribution given in [5]. If $g_0$ is the density of a $N(0, 1)$ distribution, the above characterization reads

$$X \sim N(0, 1) \iff E[Xf'_0(X) + (1 - X^2)f_0(X)] = 0 \quad (3.9)$$

for all (differentiable) $f_0 \in \mathcal{F}_0$. There is more to be said on the potential applications of (3.9). We defer this discussion to Section 4.

3.4. Discrete characterizations

Our last general result concerns discrete distributions. In this instance there is, in general, no unique interpretation of the parameters of interest; it depends on the law under investigation. As will be clear from the proof of Corollary 3.3 below (see the Appendix), our approach in this setting allows us to dispense with Assumption $B$, which was needed in order to ensure Condition (i) in Definition 2.1. However we need to strengthen Assumption $A$ as follows.

Assumption $A'$ : for $\psi(x; \theta) := \partial_u (g(x; u)/g(0; u))|_{u=\theta}$, there exists a neighborhood $\Theta_0$ of $\theta_0$ and a summable function $h : \mathbb{Z} \to \mathbb{R}^+$ such that

$$\left| \Delta^+_x \left( \frac{\psi(x; \theta)}{\psi(x; \theta_0)} \sum_{j=0}^{x-1} l_A(j; \theta_0, \theta_0)g(j; \theta_0) \right) \right| \leq h(x).$$
over $X$ for all $\theta \in \Theta_0$ and for all $A \in B_X$, where $l_A(j; \theta_0, \theta_0)$ is defined as in the proof of Theorem 2.1 and where $\Delta^+_n$ is the forward difference with respect to $x$.

Assumption $A'$ is sufficient to ensure Condition (iii) in the discrete setting. It is not restrictive and is satisfied by all the (discrete) distributions we have considered. For example, in the Poisson case, the ratio $\psi(x; \theta)/\psi(x; \theta_0)$ is none other than $(\lambda/\lambda_0)^{x-1}I_{\Theta_0}(x)$ so that known arguments (see page 65 of [14]) apply.

**Corollary 3.3.** Let $k = p = 1, X = \mathbb{Z}$ and $\Theta \subset \mathbb{R}$, and fix $\theta_0 \in \Theta$. Define $G_{\text{dis}}$ as the collection of $\theta$-parametric discrete densities $g(\cdot; \theta) : X \to [0, 1]$ with support $S \subset X$, which we take of the form $S = [N] := \{0, \ldots, N\}$ for some $N \in \mathbb{N}_0 \cup \{\infty\}$ not depending on $\theta$, such that $g \in G$ and satisfies Assumption $A'$ at $\theta_0$. Define $F_0$ as the collection of all functions $f_0 : X \to \mathbb{R}$ for which there exists a summable function $h : \mathbb{Z} \to \mathbb{R}^+$ such that $|\Delta^+_n(f_0(x))\partial_{\theta_0}(g(x; \theta_0)/g(0; \theta_0))|_{\theta=\theta_0} \leq h(x)$ over $X$ for all $\theta \in \Theta_0$, with $\Theta_0$ as in Assumption $A'$.

Then $F_0$ is $\theta$-characterizing for $g$ at $\theta_0$, with $\theta$-characterizing operator

$$T_{\theta_0}(f_0, g)(x) = \frac{\Delta^+_n(f_0(x)) \partial_{\theta_0}(g(x; \theta_0)/g(0; \theta_0))|_{\theta=\theta_0}}{g(x; \theta_0)}.$$

Corollary 3.3 contains a number of well-known discrete characterizations covered in the literature among which, for instance, those for the Poisson (see the operator in Example 2.2), the geometric $\text{Geom}(p)$, with $p$-characterizing operator

$$T_p(f_0, g)(x) = -\frac{1}{p} \left((x + 1)f_0(x + 1) - \frac{x}{1 - p}f_0(x)\right)I_0(x),$$

or the binomial $\text{Bin}(n, p)$, with $p$-characterizing operator

$$T_p(f_0, g)(x) = (1 - p)^{-n-2} \left((n - x)f_0(x + 1) - \frac{1 - p}{p}xf_0(x)\right)I_{[n]}(x).$$

The same arguments allow, of course, for dealing with other perhaps more exotic discrete distributions. Consider, for the sake of illustration, the case of the multinomial $M(n, p_1, \ldots, p_k)$, with density $g(x)$ given by

$$g(x) = \frac{n!}{\prod_{j=0}^k x_j!} \prod_{j=0}^k p_j^{x_j} I_{\Delta^n}(x) \quad (3.10)$$

where $x_0 = n - \sum_{j=1}^k x_j, p_0 = 1 - \sum_{j=1}^k p_j$ and

$$\Delta^n = \{(x_1, \ldots, x_k) \in \mathbb{N}^k \mid 0 \leq x_1 + \ldots + x_k \leq n\}.$$

In the same spirit as our previous multivariate characterizations, we start by transforming the problem into a univariate one. For this choose $p_1$ to be the parameter of interest, and rewrite (3.10) as

$$g(x) = \left(\frac{\tilde{n}_1}{x_1}\right) p_1^{x_1} (\tilde{p}_1 - p_1)^{n_1 - x_1} \frac{n! \prod_{j=2}^k p_j^{x_j}}{\tilde{n}_1 \prod_{j=2}^k x_j!} I_{\Delta^n}(x)$$
where, letting \( \bar{x}_1 = \sum_{j=2}^{k} x_j, \) we denote \( \bar{n}_1 = n - \bar{x}_1 \) and \( \bar{p}_1 = 1 - \sum_{j=2}^{k} p_j \).

Straightforward computations readily yield the corresponding operator

\[
T_{\nu}(f_0, g)(x) = \xi(x; n) \left( (\bar{n}_1 - x_1)f_0(x_1 + 1) - \frac{\bar{p}_1 - p_1}{p_1} x_1 f_0(x_1) \right) I_{\Delta^s}(x),
\]

with

\[
\xi(x; n) = \frac{\bar{p}_1}{(p_1 - \bar{p}_1)^{n_1 + 2}}.
\]

In each of the above cases, determining sufficient conditions on the test functions \( f_0 \) for the operators to be \( \theta \)-characterizing is now a simple exercise which is left to the reader.

### 3.5. Two particular cases

In this final section, we tackle two examples which do not fall within the scope of the previous general results.

First take the target distribution \( g \) to be the density of a uniform \( U[a, b] \) for \( a \leq b \in \mathbb{R} \), and define \( a \) to be the parameter of interest. This law is not, stricto sensu, a member of the scale family. It is, however, easily seen that it belongs to \( \mathcal{G} \) for all \( a \neq b \) and satisfies Assumptions A and B at all \( a < b \), with \( b \) fixed. It is readily seen that the exchanging operator \( T(f_0; a)(x) = (x - b)/(b - a)f_0((x - a)/(b - a)) \) yields the precious relationship (3.1), with \( T(f_0; a) = f_0((x - a)/(b - a)) \). This leads to the following result (the proof is left to the reader).

**Corollary 3.4.** Let \( F_0 \) be the collection of all functions \( f_0 : \mathbb{R} \to \mathbb{R} \) which are differentiable (in the sense of distributions) on \([0, 1]\). Then \( F_0 \) is \( a \)-characterizing for \( g \), with \( a \)-characterizing operator

\[
T_a(f, g)(x) = \frac{1}{b - a} \left( \frac{x - b}{b - a} f_0 \left( \frac{x - a}{b - a} \right) + f_0 \left( \frac{x - a}{b - a} \right) \right) I_{[a,b]}(x) - f_0(0)
\]

for all \( f_0 \in F_0 \).

Similarly, one can also construct a \( b \)-characterizing operator and a \( b \)-characterization for the uniform law on \([a, b]\). A third way to characterize this law is to proceed as in Section 3.2 and construct a \( \mu \)-characterization, for \( \mu \) a location parameter introduced by considering the density \( g(x - \mu) \) and working, through Corollary 3.1, with respect to \( \mu \) (see [33] for an explicit expression).

Secondly take the target distribution \( g \) to be the density of a Student \( T(\nu) \) with parameter of interest \( \nu \in \mathbb{R}_0^+ \), the tail weight parameter. This law belongs to \( \mathcal{G} \) for all \( \nu > 0 \) and satisfies Assumption A at all \( \nu > 0 \). It is readily seen that the exchanging operator

\[
T(f_0; \nu)(x) = -\frac{1}{2\nu} \frac{\Gamma(\nu/2)}{\Gamma(\nu + 1)/2} x \left( 1 + \frac{x^2}{\nu} \right)^{-\nu/2} f_0 \left( \frac{x^2}{\nu} \right)
\]
yields the precious relationship (3.1), with
\[ \hat{T}(f_0; \nu)(x) = \frac{\Gamma(\nu/2)}{\Gamma((\nu + 1)/2)} (1 + x^2/\nu)^{\nu/2} f_0(x^2/\nu). \]

Sufficient conditions on \( f_0 \) for the now usual requirements to be fulfilled are easily imposed. This leads to the following result.

**Corollary 3.5.** Fix \( \nu > 2 \). Let \( F_0 \) be the collection of differentiable (in the sense of distributions) functions \( f_0 : \mathbb{R} \to \mathbb{R} \) such that \( |f_0(x^2)|/\sqrt{1 + x^2} \) and \( |x f_0'(x^2)| \) are \( \mathbb{R} \)-integrable. Then \( F_0 \) is \( \nu \)-characterizing for \( g \), with \( \nu \)-characterizing operator
\[ T_\nu(f_0, g)(x) = \xi(x; \nu) \left( 2x^2 f_0' \left( \frac{x^2}{\nu} \right) - f_0 \left( \frac{x^2}{\nu} \right) \left( \frac{x^2}{1 + x^2} - \nu \right) \right), \]
where \( \xi(x; \nu) = -\Gamma(\nu/2)(2\nu^2\Gamma((\nu + 1)/2))^{-1} (1 + x^2/\nu)^{\nu/2}. \)

The proof of this result is mainly computational and follows along the same lines as that of all other similar results provided in this paper.

It seems appropriate to conclude on this final example. Obviously, similar parameter-based characterizations can be obtained, by means of the same tools, for gamma, hypergeometric, Laplace, Pareto distributions, etc.

**4. Further work**

In all works related with the Chen-Stein method, the characterization is merely the first (and simplest) step in a complicated process. After identifying such a characterization, the mechanics behind the Chen-Stein method can be summarized as follows.

Suppose that, for a given target distribution \( g \), we dispose of characterizations of the form \( Z \sim g(\cdot; \theta_0) \iff \mathbb{E}[T_{\theta_0}(f, g)(Z)] = 0 \) for all \( f \) in \( \mathcal{F}(g; \theta_0) \), where \( T_{\theta_0}(f, g) \) is as defined in (2.1) or as in all subsequent results. Now suppose that we are interested in studying a random object \( W \) whose distribution we do not know but which we believe to be approximately that of \( Z \). After choosing a metric \( d_H(W, Z) = \sup_{h \in \mathcal{H}} |\mathbb{E}[h(W) - h(Z)]| \) for our approximation (where \( \mathcal{H} \) is a certain class of functions, leading to, e.g., the Wasserstein metric, the Kolmogorov metric, the Total Variation metric, etc.), one can use the characterization to write, for all \( h \in \mathcal{H} \),
\[ d_H(W, Z) = \sup_{h \in \mathcal{H}} |\mathbb{E}[T_{\theta_0}(f_h, g)(W)]| \tag{4.1} \]
with \( f_h \) the solution of the Chen-Stein equation
\[ T_{\theta_0}(f_h, g)(x) = h(x) - \mathbb{E}[h(Z)]. \tag{4.2} \]

The intuitive argument behind the method is that if the law of \( W \) is close to that of \( Z \), then the rhs of (4.1) should be close to 0. Starting from there,
the usual methodology relies on three steps, namely (i) solving (4.2) for all $h \in \mathcal{H}$, (ii) deriving bounds – the so-called magic factors – on the corresponding solutions, and (iii) applying the right tool (exchangeable pairs, zero- or size-biased distributions, truncation, etc.) in order to obtain explicit bounds on the right-hand side of (4.1) through the bounds obtained in step (ii). We refer the reader to [4] and [10] or to (nearly) any of the references given at the end of this paper for a more complete overview of the method.

A careful reading of the different proofs provided in this paper shows that Theorem 2.1 not only yields Chen-Stein operators, but also solutions to the corresponding Stein equations (see equations (2.3), (A.1) and (A.4)). Hence step (i) above is provided for by our result. This, however, is only the tip of the iceberg, since steps (ii) and (iii) are the ones which require the most work, and at this point we do not yet know if there exists a general approach to the method by means of our parametric characterizations which would allow to obtain some form of universal bounds. We nevertheless believe that there exist a number of promising applications, which we will now briefly outline.

As already mentioned previously, an interesting direction of research lies in studying how a combination of different characterizations for any specific target distribution could allow for improvements on known results. This view is supported by a recent success, reported in [5], where a crafty combination of a location-based and a scale-based characterization for the exponential distribution allowed to derive good approximation bounds in problems of exponential convergence. In the same spirit one could perhaps combine, for instance in normal approximation theorems, the usual (location) characterization with its scale counterpart (3.9) for some specific problems in which the former alone does not yield satisfactory results. This is the subject of ongoing research.

A second example of interesting connection concerns the so-called exchangeable pairs method (see, for instance, [2] for an introduction). Although we dispose, at this time, of no general result, we can readily show the following. Construct an exchangeable pair $W, W'$ such that $E[W'|W] = (1 - 2/\lambda_1)W + r_1(W)$ and $E[W'^2 - 1|W] = (1 - 2/\lambda_2)W^2 + r_2(W)$, where $r_1$ and $r_2$ are remainder terms. (For instance, when $W$ is a normalized partial sum of $n$ i.i.d. random variables $\xi_i$, the usual construction applies, yielding $\lambda_1 = \lambda_2 = n$ and $r_1(W) = 0$, $r_2(W) = (c_n - 1)/n$ with $c_n = E[\sum \xi_i^2]$.) Standard results for exchangeable pairs show that (3.9) can be rewritten as

$$E[Wf'(W) - (W^2 - 1)f(W)] = \lambda_1 E[(W - W')(f'(W) - f'(W'))]$$
$$- \lambda_2 E[(W^2 - W'^2)(f(W) - f(W'))] + r_f(W),$$

where $r_f$ is a remainder term. An approach similar to that of [5] could then, possibly, allow for recovering rates of convergence by means of the exchangeable pairs approach through the scale-based characterization for the normal law. More generally, when playing around with different densities (exponential, Laplace, Gumbel, etc.) and our characterization theorems, there appears to be a general pattern in the construction of an exchangeable pair for a given problem.
of convergence towards a law $g$ in terms of the behavior of the operator $\mathcal{T}_\theta$. This is the subject of ongoing research.

Thirdly, regarding the existing characterizations for the multivariate Gaussian given in, e.g., [30] or [6], one sees that, although these share similarities with (3.6) and (3.7), they require the presence of second derivatives of the test functions $f$. It is perhaps intriguing to note that the same smoothness restrictions occur in the univariate case where characterizations obtained through the generator approach also require higher order derivatives than those obtained through the density approach (see [29]). As argued in [6], the presence of second derivatives imposes restrictive smoothness conditions. In view of (3.6) and (3.7), such restrictions appear to be non-structural. It would be interesting to see if any advantage can be gained on existing results through our characterizations. More generally, we have by no means fully explored the possibilities of our parametric approach to multivariate characterizations; it is possible that more skillful manipulations will yield characterizations better adapted to stochastic approximation theorems.

Finally, there appears to be a connection between our characterizations and the family of orthogonal polynomials and distributional transformations described in [18] or [19]. For the sake of illustration we mention one such – perhaps anecdotal – connection. Consider the sequence of Hermite polynomials given by $H_0(x) = 1, H_1(x) = x, H_2(x) = x^2 - 1, H_3(x) = x^3 - 3x$, and so on. With these notations, the usual location-based characterization for the Gaussian can be rewritten

$$X \sim N(0,1) \iff E[f'(X)H_0(X) - H_1(X)f(X)] = 0$$

for all $f$ in a sufficiently large class, and the scale-based characterization (3.9) can be rewritten

$$X \sim N(0,1) \iff E[f'(X)H_1(X) - H_2(X)f(X)] = 0$$

for all $f$ in a sufficiently large class. Moreover, when playing around with different densities (exponential, Laplace, Gumbel, etc.) and our characterization theorems, one easily sees that a similar phenomenon occurs for these laws as well, though with other types of polynomials. This begs for an interpretation in terms of the parametric properties of the target distribution.

### Appendix A: Technical proofs

**Proof of equality (2.4).** First note that

$$\partial_t (f_A(x; t)g(x; t))|_{t=\theta} = \partial_t \left( \int_{\theta_0}^t l_A(x; u, t)g(x; u)dm_\Theta(u) \right)|_{t=\theta} = l_A(x; \theta, \theta)g(x; \theta) + \int_{\theta_0}^\theta \partial_t (l_A(x; u, t))|_{t=\theta} g(x; u)dm_\Theta(u).$$

Now we have

$$\partial_t (l_A(x; u, t))|_{t=\theta} = \partial_t (\mathbb{1}_{S_\theta}(x))|_{t=\theta} = \mathbb{1}_{A}(x) - P(Z_u \in A | Z_u \in S_\theta) - \partial_t (P(Z_u \in A | Z_u \in S_\theta)) \mathbb{1}_{S_\theta}(x).$$
On the one hand, we easily see that the function

$$H_1(x; \theta) := \int_{\theta_0}^{\theta} \partial_t (1_{S_t}(x))|_{t=\theta} \int_{\theta_0}^{\theta} (1_{A}(x) - P(Z_u \in A | Z_u \in S_t)) g(x; u) dm_\theta(u)$$

is well-defined, bounded uniformly in \( \theta \) over \( \Theta_0 \) by a \( m_\mathcal{L} \)-integrable function and satisfies \( H_1(x; \theta_0) = 0 \). On the other hand, we have

$$\partial_t (P(Z_u \in A | Z_u \in S_t))|_{t=\theta} = \frac{\partial_t (P(Z_u \in A \cap S_t))|_{t=\theta}}{P(Z_u \in S_t)} - \partial_t (P(Z_u \in S_t))|_{t=\theta} \frac{P(Z_u \in A \cap S_t)}{P(Z_u \in S_t)^2},$$

where clearly both derivatives are well-defined. Hence the function

$$H_2(x; \theta) := 1_{S_\theta}(x) \int_{\theta_0}^{\theta} \partial_t (P(Z_u \in A | Z_u \in S_t))|_{t=\theta} g(x; u) dm_\theta(u)$$

is also well-defined, bounded uniformly in \( \theta \) over \( \Theta_0 \) by a \( m_\mathcal{L} \)-integrable function and satisfies \( H_2(x; \theta_0) = 0 \). Defining

$$H(x; \theta) := H_1(x; \theta) - H_2(x; \theta)$$

we see that all the assertions in the proof of Theorem 2.1 hold, and, moreover, that

$$\partial_t \left( f_A(x; t) g(x; t) \right)|_{t=\theta_0} = l_A(x; \theta_0, \theta_0) g(x; \theta_0) + H(x; \theta_0) = l_A(x; \theta_0, \theta_0) g(x; \theta_0).$$

This completes the proof of Theorem 2.1. \( \square \)

**Proof of Corollary 3.1 (location).** We apply the method described in Section 3.1.

**Step 1:** Choose \( T(f_0; \mu)(x) = f_0(x - \mu) \).

**Step 2:** Set \( T(f_0; \mu)(x) = - f_0(x - \mu) \).

**Step 3:** One easily sees that, for any \( f_0 \in \mathcal{F}_0 \), Conditions (\( \mu \)-i)-(\( \mu \)-iii) on \( f_0 \) entail that Conditions (i)-(iii) are satisfied by \( T(f_0; \mu)(x) \).

**Step 4:** Consider the solution of the Chen-Stein equation given by

$$f_A^1(x - \mu_0) = - \frac{1}{g_0(x - \mu_0)} \left( \int_{x_0}^{x} l_A(y; \mu_0, \mu_0) g_0(y - \mu_0) dm \chi(y) + c(x) \right)$$

for some \( x_0 \in \mathcal{X} \), where the function \( x \mapsto c(x) \) has derivative (in the sense of distributions) equal to zero and is defined in such a way that (\( \int_{x_0}^{x} l_A(y; \mu_0, \mu_0) g_0(y - \mu_0) dm \chi(y) + c(x) \)).
\(\mu_0)dm_x(y) + c(x))\partial_x I_S(x - \mu_0) = 0\) over \(X\). This function can be expressed as a sum of Dirac delta functions whose vertices are determined by \(\partial_x I_S(x - \mu_0)\). This yields the candidate solution

\[
f_0^A(x) = -\frac{1}{g_0(x)} \left( \int_{x_0}^{x + \mu_0} l_A(y; \mu_0, \mu_0)g_0(y - \mu_0)dm_x(y) + c(x + \mu_0) \right).
\] (A.1)

For this function to belong to \(F_0\), we need Condition (\(\mu\)-ii), which is obvious, Condition (\(\mu\)-iii), which is also obvious thanks to Assumption A once again, and Condition (\(\mu\)-i) which will hold as soon as

\[
\left| \int_X f_0^A(x)g_0(x)dm_x(x) \right| = C + \left| \int_X \int_{x_0}^{x + \mu_0} l_A(y; \mu_0, \mu_0)g(y - \mu_0)dm_x(y)I_S(x)dm_x(x) \right| < \infty,
\]

where \(C = \int_X c(x + \mu_0)I_S(x)dm_x(x)\) is finite. Since Assumption B then ensures that the quantity \(\int_X f_0^A(x)g_0(x)dm_x(x)\) is bounded, Condition (\(\mu\)-i) is satisfied as well, which concludes the proof.

**Proof of Corollary 3.3 (discrete).** In this framework, the exchangeability condition (3.1) reads

\[
\partial_0(\tilde{T}(f_0; \theta)(x)g(x; \theta))|_{\theta = \theta_0} = \Delta_+^+(T(f_0; \theta_0)(x)g(x; \theta_0)),
\] (A.2)

for some \(f_0 \in F_0\). In order to obtain the announced \(\theta\)-characterizing operator \(T_{\theta_0}(f_0, g)\), we define

\[
\tilde{T}(f_0; \theta)(x) = \frac{\Delta_+^+(f_0(x)g(x; \theta))}{g(x; \theta)g(0; \theta)},
\] (A.3)

and the (invertible) exchanging operator

\[
T(f_0; \theta_0)(x) = f_0(x)\frac{\partial_0(\frac{g(x; \theta)g(0; \theta)}{g(x; \theta_0)})|_{\theta = \theta_0}}{g(x; \theta_0)}.
\]

One readily checks that these choices satisfy the exchangeability condition (A.2).

Fix \(\theta_0 \in \Theta\). The sufficient condition is immediate. For the necessary condition to hold, we solve

\[
\Delta_+^+(T(f_0^A; \theta_0)(x)g(x; \theta_0)) = l_A(x; \theta_0, \theta_0)g(x; \theta_0),
\]

with \(l_A\) as before, to obtain the candidate solution

\[
f_0^A(x) = (\psi(x; \theta_0))^{-1} \sum_{j=0}^{x-1} l_A(j; \theta_0, \theta_0)g(j; \theta_0),
\] (A.4)

where the sum over an empty set is 0. Assumption A’ guarantees that this function belongs to \(F_0\).
Acknowledgements

We especially wish to thank Marc Hallin for the initial inspiration behind this work. We also thank our colleague Baptiste Chantraine, as well as all the participants at the ULB-VUB seminar where this result was first presented, for their comments and suggestions which led to substantial improvements.

References


