Annals of the Institute of Statistical Mathematics (2021) 73:31–59
https://doi.org/10.1007/s10463-019-00735-1

Fixed point characterizations of continuous univariate probability distributions and their applications

Steffen Betsch[1] · Bruno Ebner[1]

Received: 11 December 2018 / Revised: 17 September 2019 / Published online: 20 November 2019

Abstract
By extrapolating the explicit formula of the zero-bias distribution occurring in the context of Stein's method, we construct characterization identities for a large class of absolutely continuous univariate distributions. Instead of trying to derive characterizing distributional transformations that inherit certain structures for the use in further theoretic endeavors, we focus on explicit representations given through a formula for the density- or distribution function. The results we establish with this ambition feature immediate applications in the area of goodness-of-fit testing. We draw up a blueprint for the construction of tests of fit that include procedures for many distributions for which little (if any) practicable tests are known. To illustrate this last point, we construct a test for the Burr Type XII distribution for which, to our knowledge, not a single test is known aside from the classical universal procedures.

Keywords Burr Type XII distribution · Density approach · Distributional characterizations · Goodness-of-fit tests · Non-normalized statistical models · Probability distributions · Stein's method

1 Introduction

Over the last decades, Stein's method for distributional approximation has become a viable tool for proving limit theorems and establishing convergence rates. At its heart

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-019-00735-1) contains supplementary material, which is available to authorized users.

✉ Steffen Betsch
Steffen.Betsch@kit.edu

Bruno Ebner
Bruno.Ebner@kit.edu

[1] Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Englerstrasse 2, 76131 Karlsruhe, Germany

lies the well-known Stein characterization which states that a real-valued random variable Z has a standard normal distribution if, and only if,

$$\mathbb{E}\big[f'(Z) - Zf(Z)\big] = 0 \tag{1}$$

holds for all functions f of a sufficiently large class of test functions. To exploit this characterization for testing the hypothesis

$$\mathbf{H_0} : \mathbb{P}^X \in \big\{\mathcal{N}(\mu, \sigma^2) \,|\, (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)\big\} \tag{2}$$

of normality, where $\mathbb{P}^X$ is the distribution of a real-valued random variable X, against general alternatives, Betsch and Ebner (2019b) used that (1) can be untied from the class of test functions with the help of the so-called zero-bias transformation introduced by Goldstein and Reinert (1997). To be specific, a real-valued random variable X^* is said to have the X-zero-bias distribution if

$$\mathbb{E}\big[f'(X^*)\big] = \mathbb{E}\big[Xf(X)\big]$$

holds for any of the respective test functions f. If $\mathbb{E}X = 0$ and $\mathrm{Var}(X) = 1$, the X-zero-bias distribution exists and is unique, and it has distribution function

$$T^X(t) = \mathbb{E}\big[X(X-t)\mathbb{1}\{X \le t\}\big], \quad t \in \mathbb{R}. \tag{3}$$

By (1), the standard Gaussian distribution is the unique fixed point of the transformation $\mathbb{P}^X \mapsto \mathbb{P}^{X^*}$. Thus, the distribution of X is standard normal if, and only if,

$$T^X = F_X, \tag{4}$$

where F_X is the distribution function of X. In the spirit of characterization-based goodness-of-fit tests, an idea introduced by Linnik (1962), this fixed point property directly admits a new class of testing procedures as follows. Letting $\widehat{T}_n^X$ be an empirical version of T^X and $\widehat{F}_n$ the empirical distribution function, both based on the standardized sample, Betsch and Ebner (2019b) proposed a test for (2) based on the statistic

$$G_n = n \int_{\mathbb{R}} \Big|\widehat{T}_n^X(t) - \widehat{F}_n(t)\Big|^2 \, w(t)\, \mathrm{d}t,$$

where w is an appropriate weight function, which, in view of (4), rejects the normality hypothesis for large values of G_n. As these tests have several desirable properties such as consistency against general alternatives, and since they show a very promising performance in simulations, we devote this work to the question to what extent the fixed point property and the class of goodness-of-fit procedures may be generalized to other distributions.

Naturally, interest in applying Stein's method to other distributions has already grown and delivered some corresponding results. Characterizations like (1) have been established en mass. [For an overview on characterizing Stein operators and further references, we recommend the work by Ley et al. (2017).] Charles Stein himself presented some ideas fundamental to the so-called density approach (see Stein 1986, Chapter VI; Stein et al. 2004, Section 5) which we shall use as the basis of our considerations. Related results for the special case of exponential families were already given by Hudson (1978) and Prakasa Rao (1979). Another approach pioneered by Barbour (1990) (see also Götze 1991) includes working with the generator of the semigroup of operators corresponding to a Markov process whose stationary distribution is the one in consideration. A third advance is based on fixed point properties of probability transformations like the zero-bias transformation. Very general distributional transformations were introduced by Goldstein and Reinert (2005) and refined by Döbler (2017). In the latter contribution, the transformations, and with them the explicit formulae, rely heavily on sign changes of the so-called biasing functions. These sign changes, in turn, depend on the parameters of the distribution in consideration which renders the explicit representations impractical for the use in goodness-of-fit testing.

The starting point of the present paper is the density approach identity. Here, a result more general than (1) is provided by showing that, for suitable density functions p, a given real-valued random variable X has density p if, and only if,

$$\mathbb{E}\left[f'(X) + \frac{p'(X)}{p(X)} f(X)\right] = 0 \tag{5}$$

holds for a sufficiently large class of test functions. We provide fixed point characterizations like (4) by using the analogy between (5) and (1) to extrapolate the explicit formula (3) of the zero-bias transformation to other distributions. Using this approach, these transformations will no longer be probability transformations, but we maintain the characterizing identity which suffices for the use in goodness-of-fit testing. Our confidence in the approach is manifested by the fact that it has already been implemented by Betsch and Ebner (2019a) for the special case of the gamma distribution.

With our results, we contribute to the growing amount of applications of Stein's (or the Stein–Chen) method and his characterization in the realm of statistics. Much has been done in the area of stochastic modeling, which often includes statistical methods. For instance, Fang (2014) and Reinert and Röllin (2010) (see also Barbour 1982; Barbour et al. 1989) tackle counting problems in the context of random graphs with Stein's method. The technique also led to further insights in time series- and mean field analysis, cf. Kim (2000) and Ying (2017). Braverman and Dai (2017) and Braverman et al. (2016) developed Stein's method for diffusion approximation which is used as a tool for performance analysis in the theory of queues. As for statistical research that is more relatable to our pursuits, quite a bit is known when it comes to normal approximation for maximum likelihood estimators, investigated, for instance, by Anastasiou (2018), Anastasiou and Gaunt (2019), Anastasiou and Reinert (2017) and Pinelis (2017), to name but a few contributions. Moreover, Gaunt et al. (2017) considered Chi-square approximation to study Pearson's statistic which is used for goodness-of-fit testing in classification problems. Also note that Anastasiou and Reinert (2018) apply the results of Gaunt et al. (2017) to obtain bounds to the Chi-square distribution for twice the log-likelihood ratio, the statistic used for the classical likelihood ratio test. Finally, the contributions by Chwialkowski et al. (2016) and Liu et al. (2016) aim at the goal we also pursue in Sect. 7, namely to apply Steinian characterizations to construct goodness-of-fit tests for probability distributions.

The paper at hand is organized as follows. We first introduce an appropriate setting for our considerations by stating the conditions for a density function to fit into our framework and prove identity (5) in this specific setting. We then give our characterization results, distinguishing between distributions supported by the whole real line, those with semi-bounded support and distributions with bounded support. Throughout, we give examples of density functions of different nature to show that our conditions are not restrictive, as well as to provide connections to characterizations that are already known and included in our statements. Next, we consider applications in goodness-of-fit testing and show that the proposed tests include the classical Kolmogorov–Smirnov and Cramér-von Mises procedures, as well as three modern tests considered in the literature. To illustrate the methods in the last part, we construct the first ever goodness-of-fit test specifically for the two-parameter Burr Type XII distribution and show in a simulation study that the test is sound and powerful compared to classical procedures.

2 Notation and regularity conditions

Throughout, let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and p a nonnegative density function supported by an interval $\mathrm{spt}(p) = [L, R]$, where $-\infty \le L < R \le \infty$, and with $\int_L^R p(x)\,\mathrm{d}x = 1$. Denoting by P the distribution function associated with p, we state the following regularity conditions:

(C1) The function p is continuous and positive on (L, R), and there exists a partition $L < y_1 < \ldots < y_m < R$ such that p is continuously differentiable on (L, y_1), $(y_\ell, y_{\ell+1})$, $\ell \in \{1, \ldots, m-1\}$ and (y_m, R).

Whenever (C1) holds, we write $S(p) = (L, R) \setminus \{y_1, \ldots, y_m\}$.

(C2) For the map $S(p) \ni x \mapsto \kappa_p(x) = \left| \frac{p'(x) \min\{P(x), 1-P(x)\}}{p^2(x)} \right|$, we have that $\sup_{x \in S(p)} \kappa_p(x) < \infty$,

(C3) $\int_{S(p)} \big(1 + |x|\big)\big|p'(x)\big|\mathrm{d}x < \infty$,

(C4) $\lim_{x \searrow L} \frac{P(x)}{p(x)} = 0$, and

(C5) $\lim_{x \nearrow R} \frac{1-P(x)}{p(x)} = 0$.

The integral $\int_{S(p)}$ is understood as the sum of the integrals over the interval components in (C1). For a probability density function p that satisfies (C1), and a function $f : (L, R) \to \mathbb{R}$ which is differentiable on $S(p)$ except in one point, we denote the point of non-differentiability in $S(p)$ by t_f and set

$$S(p, f) = S(p) \setminus \{t_f\} = (L, R) \setminus \{y_1, \ldots, y_m, t_f\}.$$

We index the elements of $(L, R) \setminus S(p, f)$ with $y_1^f < \ldots < y_{m+1}^f$.

Definition 1 (*Test functions*) For a probability density function p with $\mathrm{spt}(p) = [L, R]$ that satisfies (C1), we denote by $\mathcal{F}_p$ the set of all functions $f : (L, R) \to \mathbb{R}$ that are continuous on (L, R) and differentiable on $S(p)$ except in (precisely) one point, that satisfy

$$\lim_{x \searrow L} f(x)\,p(x) = \lim_{x \nearrow R} f(x)\,p(x) = 0,$$

and for which $x \mapsto \frac{p'(x)}{p(x)} f(x)$ and $x \mapsto f'(x)$ are bounded on $S(p, f)$.

We write $\mathcal{L}^1$ for the Borel–Lebesgue measure on the real line, and $X \sim p\mathcal{L}^1$ whenever a random variable X has Lebesgue density p, and we define

$$\mathrm{disc}(X) = \big\{t \in (L, R) \,\big|\, \mathbb{P}(X = t) > 0\big\},$$

the set of all atoms of a random variable X, containing at most countably many points.

3 The density approach identity

In this section, we restate the density approach identity. Since we use a very particular class of test functions, which bears some technicalities, we give an outline of the proof in the supplementary material, Section 1, roughly following Ley and Swan (2013b). We refer to Section II of Ley and Swan (2013a) for a discrete version of the density approach identity and mention Ley and Swan (2016) for related statements in the context of parametric distributions.

Lemma 1 *If p is a probability density function with $\mathrm{spt}(p) = [L, R]$ that satisfies (C1) and (C2), and if $X : \Omega \to (L, R)$ is a random variable with $\mathbb{P}\big(X \in S(p)\big) = 1$, then $X \sim p\mathcal{L}^1$ if, and only if,*

$$\mathbb{E}\left[f'(X) + \frac{p'(X)}{p(X)}\, f(X)\right] = 0$$

for each $f \in \mathcal{F}_p$ with $t_f \notin \mathrm{disc}(X)$.

Remark 1 Note that some contributions to the scientific literature (like Ley and Swan 2011, 2013b; Betsch and Ebner 2019a) claim that the function

$$(L, R) \ni x \mapsto \int_L^x \Big(\mathbb{1}_{(L,t]}(s) - P(t)\Big) p(s)\, \mathrm{d}s$$

is differentiable when, in fact, it fails to be so in exactly one point, namely in t. This leads to the unfortunate consequence that we cannot assume functions in $\mathcal{F}_p$ to be differentiable, and if the random variable X is discrete with an atom at the point of non-differentiability of a test function, the expectation in Lemma 1 makes no sense. As such, the error has no consequence for Ley and Swan (2013b), since they only consider absolutely continuous random variables for which $\mathrm{disc}(X) = \emptyset$. For the general case, the restriction to test functions with $t_f \notin \mathrm{disc}(X)$ becomes necessary.

Remark 2 Since $f_t^{p\,\prime}(x) + \frac{p'(x)}{p(x)} f_t^p(x)$ is uniformly bounded over $x \in S(p, f_t^p)$ by Equation (1.1) in the supplementary material, we can assume that, for each $f \in \mathcal{F}_p$, the function $f' + \frac{p'}{p} f$ is integrable with respect to any probability measure $\mathbb{P}^X$ such that $X \in S(p, f)$ $\mathbb{P}$-almost surely. Note that conditions comparable to our assumptions (C1)–(C5) are commonly stated in the context of Stein's method [see, e.g., Section 13 by Chen et al. (2011), Section 4 by Chatterjee and Shao (2011), or Döbler (2015)]. See also Remark 7 for further comments on the regularity conditions. It is easy to adapt the proof of Lemma 1 so that we can also allow for finitely many points in which the density function is zero [by changing condition (C1) accordingly]. However, for our characterization results later on we need the continuity of the functions in $\mathcal{F}_p$ on the whole interval (L, R), and this we cannot get from f_t^p when we allow for zeros in the function p.

Remark 3 For later use, we note that if (C4) holds, any function $f \in \mathcal{F}_p$ is subject to $\lim_{x \searrow L} f(x) = 0$, since f_t^p from the proof of Lemma 1 satisfies

$$\lim_{x \searrow L} f_t^p(x) = \lim_{x \searrow L} \frac{P(x)}{p(x)} \big(1 - P(t)\big) = 0.$$

By analogy, if (C5) holds, each function $f \in \mathcal{F}_p$ can be taken to satisfy $\lim_{x \nearrow R} f(x) = 0$.

In a different form, the characterization given in Lemma 1 has successfully been applied for distributional approximations in the Curie–Weiss model (see Chatterjee and Shao 2011) or the hitting times of Markov chains (see Peköz and Röllin 2011). For an overview, we refer to Section 13 by Chen et al. (2011). In this paper, however, we use the characterization to derive another, more explicit identity that typifies distributions with density functions as above. We thereby generalize the fixed point properties of the well-known zero-bias and equilibrium transformations but also classical identities, such as the characterization of the exponential distribution through the mean residual life function.

4 Univariate distributions supported by the real line

Assume for now that $p : \mathbb{R} \to [0, \infty)$ is a probability density function supported by the whole real line.

Theorem 1 *Suppose that p is a probability density function with* $\mathrm{spt}(p) = \mathbb{R}$ *that satisfies (C1)–(C3). Let $X : \Omega \to \mathbb{R}$ be a random variable such that* $\mathbb{P}\big(X \in S(p)\big) = 1$, *and*

$$\mathbb{E}\left|\frac{p'(X)}{p(X)}\right| < \infty, \quad \mathbb{E}\left|\frac{p'(X)}{p(X)}\, X\right| < \infty. \tag{6}$$

Then, $X \sim p\mathcal{L}^1$ if, and only if, the distribution function of X has the form

$$F_X(t) = \mathbb{E}\left[\frac{p'(X)}{p(X)}\,(t - X)\,\mathbb{1}\{X \le t\}\right], \quad t \in \mathbb{R}.$$

The proof is given in the supplementary material, Section 2.

Remark 4 For a density function p supported by the whole real line which satisfies (C1)–(C3), take the set of all distributions considered in Theorem 1, that is,

$$\mathcal{P} = \left\{ \mathbb{P}^X \;\middle|\; \mathbb{P}\big(X \in S(p)\big) = 1,\ \mathbb{E}\left|\frac{p'(X)}{p(X)}\right| < \infty,\ \text{and } \mathbb{E}\left|\frac{p'(X)}{p(X)}\, X\right| < \infty \right\}.$$

The previous theorem concerns properties of the mapping

$$\mathcal{T}\colon \mathcal{P} \to D(\mathbb{R}), \quad F_X \mapsto \mathcal{T}(F_X) = \left(t \mapsto \mathbb{E}\left[\frac{p'(X)}{p(X)} \,(t - X)\, \mathbb{1}\{X \le t\} \right] \right),$$

where $D(\mathbb{R})$ is the càdlàg space over $\mathbb{R}$, and where we identified elements from $\mathcal{P}$ with their distribution function. In particular, Theorem 1 states that this mapping has a unique fixed point, namely $\mathbb{P}^X = p\mathcal{L}^1$. Putting further restrictions on the distribution of X such that d_p^X from the proof of Theorem 1 (see supplementary material, Section 2) is a probability density function without assuming that F_X is given through our explicit formula, we have actually shown in the last calculation of that proof the existence of a distribution for some random variable X_p with

$$\mathbb{E}\big[f'(X_p)\big] = \mathbb{E}\left[-\frac{p'(X)}{p(X)}\, f(X) \right]$$

for each $f \in \mathcal{F}_p$, and we could think of $\mathcal{T}$ as a distributional transformation. These additional restrictions [for the normal distribution, they are $\mathbb{E}X = 0$ and $\mathrm{Var}(X) = \sigma^2$, see Example 1] scale down the class of distributions in which the characterization holds. Therefore, our point is not to cling on to distributional transformations, which makes explicit formulae more complicated [as witnessed by Döbler (2017), Remark 1(d) and Remark 2], but extract whichever information we can get from the explicit formula itself.

In the proof of Theorem 1, we have actually also shown another characterization result, but via the density function.

Corollary 1 *Let p be a probability density function with $\mathrm{spt}(p) = \mathbb{R}$ that satisfies (C1)–(C3). When $X : \Omega \to \mathbb{R}$ is a random variable with density function $\mathfrak{f}_X$, $\mathbb{E}\left|\frac{p'(X)}{p(X)}\right| < \infty$, and $\mathbb{E}\left|\frac{p'(X)}{p(X)}\, X\right| < \infty$, then $X \sim p\mathcal{L}^1$ if, and only if, the density function of X has the form*

$$\mathfrak{f}_X(t) = \mathbb{E}\left[\frac{p'(X)}{p(X)}\, \mathbb{1}\{X \le t\} \right], \quad t \in \mathbb{R}.$$

It is clear from the proof that it suffices to have the above representation for the density function of X only for $\mathcal{L}^1$ -almost every (a.e.) $t \in \mathbb{R}$ to conclude that $X \sim p\mathcal{L}^1$. This is much in line with the intuition about density functions, since they uniquely determine a probability law, but are themselves only unique $\mathcal{L}^1$-almost everywhere (a.e.).

To get a feeling for the results, we consider two examples. For brevity, we only give the characterization via Theorem 1, the result via Corollary 1 being clear from that.

Example 1 (Mean-zero Gaussian distribution) For $x \in \mathbb{R}$, let

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{x^2}{2\sigma^2}\Big),$$

where $0 < \sigma < \infty$. The function p is positive and continuously differentiable on the whole real line, so (C1) is satisfied [with $m = 0$ and $S(p) = \mathbb{R}$]. We have $\frac{p'(x)}{p(x)} = -\frac{x}{\sigma^2}$, $x \in \mathbb{R}$. Condition (C3) follows from the existence of mean and variance of the normal distribution, and (C2) is proven using the (easily verified) identities

$$\frac{1 - P(x)}{p(x)} \leq \frac{\sigma^2}{x}, \quad x > 0, \quad \text{and} \quad \frac{P(x)}{p(x)} = \frac{1 - P(-x)}{p(-x)} \leq -\frac{\sigma^2}{x}, \quad x < 0.$$

By Theorem 1, a real-valued random variable X with $\mathbb{E}X^2 < \infty$ follows the mean-zero Gaussian law with variance σ^2 if, and only if, the distribution function of X has the form

$$F_X(t) = \mathbb{E}\left[\frac{X}{\sigma^2}(X - t)\,\mathbb{1}\{X \leq t\}\right], \quad t \in \mathbb{R}.$$

In this particular example, the map $\mathcal{T}$ introduced in Remark 4 is, up to a change of the domain, the zero-bias transformation discussed in the introduction. The transformation $\mathbb{P}^X \mapsto \mathbb{P}^{X^*}$ (using notation from our introduction), which coincides with our mapping $\mathcal{T}$ in terms of the law of the maps, has the normal distribution $\mathcal{N}(0, \sigma^2)$ as its unique fixed point and thus typifies this distribution within all distributions with mean zero and variance σ^2. The message of the example at hand is that our characterization result (Theorem 1) has the characterization via the zero-bias distribution as a special case. It is notable that we generalize this well-known characterization in the sense that the explicit formula given above identifies the normal distribution $\mathcal{N}(0, \sigma^2)$ not only within the class of all distributions with mean zero and variance σ^2, but within the class of all distributions with $\mathbb{E}X^2 < \infty$. However, if $\mathbb{E}X \neq 0$ or $\mathrm{Var}(X) \neq \sigma^2$, the formula for F_{X^*} may no longer be a distribution function, and $\mathcal{T}$ is to be understood as an extension of the operator that maps $\mathbb{P}^X \mapsto \mathbb{P}^{X^*}$ onto the larger domain

$$\mathcal{P} = \left\{\mathbb{P}^X \,\middle|\, \mathbb{E}X^2 < \infty\right\} \supsetneq \left\{\mathbb{P}^X \,\middle|\, \mathbb{E}X = 0 \text{ and } \mathrm{Var}(X) = \sigma^2\right\}.$$

The conditions (C1)–(C3) also hold for the normal distribution with location parameter included. We simply chose the setting above to illustrate the connection to the zero-bias distribution.

Example 2 (Laplace distribution) For a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$, consider the density of the corresponding Laplace distribution,

$$p(x) = \frac{1}{2\sigma} \exp\Big(-\frac{|x - \mu|}{\sigma}\Big), \quad x \in \mathbb{R}.$$

Condition (C1) is satisfied with $m = 1$, $y_1 = \mu$, and $S(p) = (-\infty, \mu) \cup (\mu, \infty)$. We have

$$\frac{p'(x)}{p(x)} = \frac{\operatorname{sign}(\mu - x)}{\sigma}, \quad x \neq \mu.$$

To verify (C2), use that the distribution function of the Laplace distribution can be given explicitly to obtain $\sup_{x \in S(p)} \kappa_p(x) \leq 1 < \infty$. Condition (C3) follows from a simple calculation. Consequently, Theorem 1 holds, and the characterization for the Laplace distribution reads as follows. A real-valued random variable X with distribution function F_X and $\mathbb{E}|X| < \infty$, which satisfies $\mathbb{P}\big(X \neq \mu\big) = 1$, has the Laplace distribution with parameters μ and σ if, and only if,

$$F_X(t) = \mathbb{E}\left[\frac{\operatorname{sign}(\mu - X)}{\sigma}\,(t - X)\,\mathbb{1}\{X \leq t\}\right], \quad t \in \mathbb{R}.$$

In the context of probability distributions on the real line, we have also checked the conditions (C1)–(C3) for the Cauchy- and Gumbel distribution, showing that we do not need any moment assumptions to prove (C3), and that the characterizations include more complicated distributions which are important in applications. We will give more examples later on.

5 Univariate distributions with semi-bounded support

In this section, we seek to provide characterization results similar to those in the previous section, but for probability distributions with semi-bounded support. We have chosen in Sect. 4 to first prove the characterization via the distribution function since this is the 'conventional' way, or at least, say, the way the special case of the zero-bias transformation is known. From a logical perspective, it is more convenient to first establish the result via the density function as in Corollary 1 and then to derive the corresponding distribution function. We first discuss the case when p is a density function whose support is bounded from below. Namely, we let $p : \mathbb{R} \to [0, \infty)$ be a probability density function with $\operatorname{spt}(p) = [L, \infty)$, $L > -\infty$. The most important case is $L = 0$, that is, density functions supported by the positive half line.

Theorem 2 *Let p be a probability density function with $\operatorname{spt}(p) = [L, \infty)$ that satisfies the conditions (C1)–(C4). If $X : \Omega \to (L, \infty)$ is a random variable with density function $\mathfrak{f}_X$, $\mathbb{E}\left|\frac{p'(X)}{p(X)}\right| < \infty$, and $\mathbb{E}\left|\frac{p'(X)}{p(X)}\,X\right| < \infty$, then $X \sim p\mathcal{L}^1$ if, and only if,*

$$\mathfrak{f}_X(t) = \mathbb{E}\left[-\frac{p'(X)}{p(X)}\,\mathbb{1}\{X > t\}\right], \quad t > L.$$

The proof of this theorem consists of arguments and calculations that are very similar to those in the proof of Theorem 1, and we refrain from giving the details. Instead, we give some insight on the special case of density functions on the positive axis.

Remark 5 The integrability condition on X can be weakened in cases where the density function p is positive and continuously differentiable, as well as supported by the positive axis, i.e., $m = 0$ and $S(p) = (0, \infty)$. In this case, the calculation in the sufficiency part of the proof of Theorem 2 reduces to

$$\begin{aligned}\mathbb{E}\big[f'(X)\big] &= \mathbb{E}\left[-\frac{p'(X)}{p(X)}\int_0^X f'(s)\,\mathrm{d}s\,\mathbb{1}\{X \le t_f\}\right] \\ &\quad + \mathbb{E}\left[-\frac{p'(X)}{p(X)}\int_0^{t_f} f'(s)\,\mathrm{d}s\,\mathbb{1}\{X > t_f\}\right] \\ &\quad + \mathbb{E}\left[-\frac{p'(X)}{p(X)}\int_{t_f}^X f'(s)\,\mathrm{d}s\,\mathbb{1}\{X > t_f\}\right] \\ &= \mathbb{E}\left[-\frac{p'(X)}{p(X)}\,f(X)\right],\end{aligned}$$

and it suffices for the use of Fubini's theorem to know that $\mathbb{E}\left|\frac{p'(X)}{p(X)}\,X\right| < \infty$. Note that this condition on X is also enough to guarantee that the expectation which defines d_p^X exists $\mathcal{L}^1$-a.e., see the supplementary material, Section 3. Consequently, it suffices to require $\int_0^\infty x|p'(x)|\,\mathrm{d}x < \infty$ instead of (C3). What is more, this last condition yields

$$\int_0^\infty \int_t^\infty \big|p'(x)\big|\,\mathrm{d}x\,\mathrm{d}t = \int_0^\infty \big|p'(x)\big| \int_0^x \mathrm{d}t\,\mathrm{d}x = \int_0^\infty x\big|p'(x)\big|\,\mathrm{d}x < \infty,$$

and thus $\int_t^\infty \big|p'(x)\big|\,\mathrm{d}x < \infty$ for $\mathcal{L}^1$-a.e. $t > 0$. This suffices to derive the necessity part of Theorem 2 with equality for $\mathcal{L}^1$-a.e. $t > 0$. Putting together these thoughts, we obtain the following special case.

Corollary 2 *(Densities supported by the positive axis) Assume that p is a probability density function with $\mathrm{spt}(p) = [0, \infty)$ that is positive and continuously differentiable on $(0, \infty)$ and satisfies (C2) and (C4). Moreover, assume that $\int_0^\infty x\big|p'(x)\big|\,\mathrm{d}x < \infty$. Let X be a positive random variable with density function $\mathfrak{f}_X$, and $\mathbb{E}\left|\frac{p'(X)}{p(X)}\,X\right| < \infty$. Then, $X \sim p\mathcal{L}^1$ if, and only if, we have for $\mathcal{L}^1$-a.e. $t > 0$ that*

$$\mathfrak{f}_X(t) = \mathbb{E}\left[-\frac{p'(X)}{p(X)}\,\mathbb{1}\{X > t\}\right].$$

Up next, we use Theorem 2 to derive a characterization result for the distribution function.

Theorem 3 *Assume that p is a probability density function supported by $[L, \infty)$ and satisfying the conditions (C1)–(C4). Let $X : \Omega \to (L, \infty)$ be a random variable with $\mathbb{P}\big(X \in S(p)\big) = 1$, $\mathbb{E}\left|\frac{p'(X)}{p(X)}\right| < \infty$, and $\mathbb{E}\left|\frac{p'(X)}{p(X)}\,X\right| < \infty$. Then, $X \sim p\mathcal{L}^1$ if, and only if,*

$$F_X(t) = \mathbb{E}\left[-\frac{p'(X)}{p(X)} \Big(\min\{X, t\} - L \Big) \right], \quad t > L.$$

The proof is given in the supplementary material, Section 3. Note that the results on the distribution function are somewhat richer than the characterizations via the density function, for the latter only identify the underlying distribution within a subset of absolutely continuous probability distributions for which a density function exists. The characterization via the distribution function does not need this restriction to absolutely continuous distributions, but only that X has no atoms in the set $(L, \infty) \setminus S(p) = \{y_1, \ldots, y_m\}$.

Remark 6 In the case where $L = 0$, and p is continuously differentiable and positive on $(0, \infty)$, Theorem 3 remains true if we replace (C3) with $\int_0^\infty x|p'(x)|\,\mathrm{d}x < \infty$, and if we further drop the first integrability condition on X and only require $\mathbb{E}\left|\frac{p'(X)}{p(X)}\,X\right| < \infty$.

We obtain the following special case of the characterization.

Corollary 3 *(Densities supported by the positive axis) Assume that p is a probability density function with* $\mathrm{spt}(p) = [0, \infty)$ *that is positive and continuously differentiable on* $(0, \infty)$ *and satisfies (C2) and (C4). Moreover, assume that* $\int_0^\infty x\left|p'(x)\right|\mathrm{d}x < \infty$. *Let X be a positive random variable with* $\mathbb{E}\left|\frac{p'(X)}{p(X)}\,X\right| < \infty$. *Then, $X \sim p\mathcal{L}^1$ if, and only if,*

$$F_X(t) = \mathbb{E}\left[-\frac{p'(X)}{p(X)} \min\{X, t\} \right], \quad t > 0.$$

Now follows the major source of examples we give in this work. We omit the explicit proofs of the regularity conditions for they consist of (sometimes) tedious calculations which provide no insight on the characterizations. Instead, we give the following remark on how the conditions are to be verified, and on their necessity in general.

Remark 7 The regularity condition (C1) is easily understood and checked for a given density function. Note that the weaker assumption of absolute continuity of p, which is mostly used in the context of Stein's method, entails similar problems as described in Remark 1: If p is merely assumed to be absolutely continuous, then in order to handle random variables X with discrete parts (e.g., in Lemma 1), we would still have to identify the points of non-differentiability of p in order to make sense of the term $p'(X)$. This would return us to considering a set like $S(p)$ which, technically, brings us to the setting we consider already.

Condition (C3) involves a direct calculation which can often be simplified if one has knowledge of the existence of moments of the distribution at hand. From the proofs of our characterizations, it is apparent that (C1) and (C3) [as well as the integrability conditions on X which are in line with (C3)] are necessary to use Fubini's theorem and the fundamental theorem of calculus. As such, we do not see any truly instrumental weaker alternative conditions (apart from the special case discussed in Remarks 5 and 6) which still rigorously allow for all calculations.

Both conditions (C4) and (C5) are trivially satisfied when the respective limit of the density function is positive, and if that is not the case, L'Hospital's rule gives a reliable handle for it. With regard to these two conditions, we refer to Proposition 3.7 of Döbler (2015) who discusses them in much detail and provides easy-to-check criteria. Moreover, this specific result from Döbler (2015) indicates strongly that the two conditions are not restrictive in practice.

To prove condition (C2), it is helpful to realize, in the case when p is continuously differentiable, that κ_p is continuous. Thus, it suffices to check that $\limsup_{x \searrow L} \kappa_p(x) < \infty$ and $\limsup_{x \nearrow R} \kappa_p(x) < \infty$ for (C2) to hold. Regularity conditions (C2) and (C4)/(C5) guarantee certain beneficial properties of the test functions from $\mathcal{F}_p$. For one, they guarantee that, for $f \in \mathcal{F}_p$, $\lim_{x \searrow L} f(x) = 0$ (or $\lim_{x \nearrow R} f(x) = 0$), see Remark 3, which we need to truly get rid of the test functions in our calculations (as in the supplementary material, Section 4). Condition (C2) is stated so that functions $f \in \mathcal{F}_p$ have uniformly bounded derivative. We use this fact in our proofs (e.g., the last calculation in the supplementary material, Section 2) to apply the fundamental theorem of calculus on f' and to justify the use of Fubini's theorem. For both arguments, the boundedness of f' is not a necessary condition, but we have not found any alternative assumption for (C2) which allows for a sound and rigorous derivation of all results.

Later on, we give an example for a distribution which fails the respective version of condition (C3) that ought to hold in order for that distribution to be included in our characterization results. For a (rather artificial) density functions which violates (C4), see Example 3.6 of Döbler (2015).

With these tools at hand, the regularity conditions for all examples below can be proven. We use Corollary 3 in each case, except for the Lévy distribution. The characterizations via the density functions are not stated explicitly to save space.

Example 3 (*Gamma distribution*) Assume that

$$p(x) = \frac{\lambda^{-k}}{\Gamma(k)}\, x^{k-1} \exp\big(-\lambda^{-1} x\big), \quad x > 0,$$

is the density function of the gamma distribution with shape parameter $k > 0$ and scale parameter $\lambda > 0$. If X is a positive random variable with $\mathbb{E}X < \infty$, then X follows the gamma law with parameters k and λ if, and only if, the distribution function of X has the form

$$F_X(t) = \mathbb{E}\left[\left(-\frac{k-1}{X} + \frac{1}{\lambda}\right) \min\{X, t\}\right], \quad t > 0.$$

Note that this result has been proven explicitly, and with a similar line of proof as our general results above, by Betsch and Ebner (2019a).

Example 4 (*Exponential distribution*) Denote the density of the exponential distribution with rate parameter $\lambda > 0$ by $p(x) = \lambda e^{-\lambda x}$, $x > 0$. This is an easy special case of the previous example, namely the gamma distribution with shape parameter $k = 1$

and scale parameter $1/\lambda$. Let X be a positive random variable with $\mathbb{E}X < \infty$. Then, X has the exponential distribution with parameter λ if, and only if,

$$F_X(t) = \lambda\, \mathbb{E}\Big[\min\{X, t\}\Big], \quad t > 0.$$

This identity is (see Baringhaus and Henze 2000) equivalent to the well-known characterization of the exponential distribution via the mean residual life function, which states that a positive random variable X with $\mathbb{E}X < \infty$ follows an exponential law if, and only if, $\mathbb{E}[X - t \,|\, X > t] = \mathbb{E}[X]$, $t > 0$. For yet another observation, assume that X is a positive random variable with $\mathbb{E}X = \lambda^{-1}$. With

$$d_p^X(t) = \mathbb{E}\left[-\frac{p'(X)}{p(X)}\,\mathbb{1}\{X > t\}\right] = \lambda\, \mathbb{P}\big(X > t\big), \quad t > 0,$$

as in the proofs of our results, we have $d_p^X \geq 0$ and

$$\int_0^\infty d_p^X(t)\,\mathrm{d}t = \lambda \int_0^\infty \mathbb{P}\big(X > t\big)\,\mathrm{d}t = \lambda\,\mathbb{E}X = 1.$$

If X^e is a random variable with density function d_p^X, the proof of Theorem 2 (see Remark 6) shows that $\mathbb{E}[f'(X^e)] = \lambda\,\mathbb{E}[f(X)]$ for each $f \in \mathcal{F}_p$. Up to a change in the class of test functions, this is the defining equation of the equilibrium distribution with respect to X. Lemma 1 implies that when restricting to $\mathbb{E}X = \lambda^{-1}$, the exponential distribution with parameter λ is the unique fixed point of the equilibrium transformation $\mathbb{P}^X \mapsto \mathbb{P}^{X^e}$. This fact is used for approximation arguments with Stein's method [see Peköz and Röllin (2011), who introduced the equilibrium distribution, as well as Chapter 13.4 by Chen et al. (2011) and Section 5 by Ross (2011)]. As in the case of the zero-bias transformation, we have generalized this characterization in the sense that the explicit formula of the equilibrium distribution uniquely identifies the exponential distribution with parameter λ within the class of all distributions $\mathbb{P}^X$ with $\mathbb{E}X < \infty$.

Example 5 (*Inverse Gaussian distribution*) Denote the inverse Gaussian density by

$$p(x) = \sqrt{\frac{\lambda}{2\pi}}\, x^{-3/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x > 0,$$

where $\mu, \lambda > 0$. If X is a positive random variable with $\mathbb{E}\big[X + X^{-1}\big] < \infty$, then X follows the inverse Gaussian law with parameters μ and λ if, and only if,

$$F_X(t) = \mathbb{E}\left[\left(-\frac{\lambda}{2X^2} + \frac{3}{2X} + \frac{\lambda}{2\mu^2}\right)\min\{X, t\}\right], \quad t > 0.$$

Now, we handle distributions that are of interest for applications. The Weibull distribution is applied in hydrology and wind speed analysis, see Singh (1987) and Carrillo et al. (2014), the Burr distribution is commonly taken as a model for household income, see Singh and Maddala (1976), and the Rice distribution appears in signal processing to describe how cancelation phenomena affect radio signals [cf. Chapter

13 of Proakis and Salehi (2008)]. The last example we give is the Lévy distribution which is used to model the length of paths that are followed by photons after reflection from a turbid media, see Section 3 of Rogers (2008). There, we provide insight on the handling of an additional location parameter which is often added to probability distributions.

Example 6 (*Weibull distribution*) For $k, \lambda > 0$, let

$$p(x) = \frac{k}{\lambda^k} \, x^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^k \right), \quad x > 0,$$

be the density function of the Weibull distribution in its usual parametrization. Let X be any positive random variable with $\mathbb{E}X^k < \infty$. Then, X has the Weibull distribution with parameters k and λ if, and only if,

$$F_X(t) = \mathbb{E}\left[\left(\frac{k\, X^{k-1}}{\lambda^k} - \frac{k-1}{X} \right) \min\{X, t\} \right], \quad t > 0.$$

Example 7 (*Burr distribution*) The Burr Type XII distribution with parameters $c, k > 0$ and scale parameter $\sigma > 0$ has density function

$$p(x) = \frac{c\,k}{\sigma} \left(\frac{x}{\sigma}\right)^{c-1} \left(1 + \left(\frac{x}{\sigma}\right)^c \right)^{-k-1}, \quad x > 0.$$

A positive random variable X has the Burr distribution with parameters $c, k, \sigma > 0$ if, and only if, the distribution function of X has the form

$$F_X(t) = \mathbb{E}\left[\left(c\,(k+1)\, \frac{X^{c-1}}{\sigma^c + X^c} - \frac{c-1}{X} \right) \min\{X, t\} \right], \quad t > 0.$$

Particularly interesting about this example is that, even though the Burr distribution is substantially more complicated than many of our other examples, no moment condition is needed for the characterization to hold, since

$$\mathbb{E}\left| \frac{p'(X)}{p(X)}\, X \right| \leq |c-1| + c\,(k+1)\, \mathbb{E}\left[\frac{X^c}{\sigma^c + X^c} \right] \leq |c-1| + c\,(k+1) < \infty.$$

This implies that the characterization is universal in the sense that it identifies the Burr distribution within the set of all probability laws on the positive axis.

Example 8 (*Rice distribution*) For parameters $k, \varrho > 0$, the density function of the Rice distribution is given by

$$p(x) = \frac{2\,(k+1)\,x}{\varrho} \exp\left(-k - \frac{(k+1)\,x^2}{\varrho} \right) I_0\left(2 \sqrt{\frac{k\,(k+1)}{\varrho}}\, x \right), \quad x > 0,$$

where I_α denotes the modified Bessel function of first kind of order $\alpha \in \mathbb{Z}$. We chose the parametrization for p that is mostly used in signal processing and is easily found under the keyword of Rician fading. Let X be a positive random variable with $\mathbb{E}X^2 < \infty$. Then, X has the Rice distribution with parameters k and ϱ if, and only if

$$F_X(t) = \mathbb{E}\left[\left(-\frac{1}{X} + \frac{2\,(k+1)\,X}{\varrho} - 2\sqrt{\frac{k\,(k+1)}{\varrho}} \cdot \frac{I_1\left(2\sqrt{\frac{k\,(k+1)}{\varrho}}\,X\right)}{I_0\left(2\sqrt{\frac{k\,(k+1)}{\varrho}}\,X\right)}\right)\min\{X, t\}\right],$$

for $t > 0$. Note that despite the complexity of the term $\frac{p'(x)}{p(x)}$, the integrability condition is $\mathbb{E}X^2 < \infty$, since the quotient of the Bessel functions cancels via $\frac{I_1(y)}{I_0(y)} \le 1$, $y > 0$.

Example 9 (*Lévy distribution*) Take $\mu \in \mathbb{R}$ and $\sigma > 0$. Let

$$p(x) = \sqrt{\frac{\sigma}{2\pi}}\,\big(x - \mu\big)^{-3/2} \exp\left(-\frac{\sigma}{2(x - \mu)}\right), \quad x > \mu,$$

denote the density function of the Lévy distribution with location parameter μ and scale parameter σ. Let X be a random variable which takes values in (μ, ∞) almost surely such that $\mathbb{E}[(X - \mu)^{-1}] < \infty$ and $\mathbb{E}[(X - \mu)^{-2}] < \infty$. Then, X has the Lévy distribution with parameters μ and σ if, and only if, the distribution function of X has the form

$$F_X(t) = \frac{1}{2}\,\mathbb{E}\left[\left(\frac{3}{X - \mu} - \frac{\sigma}{(X - \mu)^2}\right)\Big(\min\{X, t\} - \mu\Big)\right], \quad t > \mu.$$

The following example is one which fails the regularity condition (C3). Recall that for distributions which are not supported by the positive axis, we need (C3) fully, that is, we cannot apply Remarks 5 or 6.

Example 10 (*Shifted gamma distribution*) Assume that

$$p(x) = \frac{\lambda^{-k}}{\Gamma(k)}\,(x - \mu)^{k-1}\,\exp\big(-\lambda^{-1}(x - \mu)\big), \quad x > \mu,$$

is the density function of the shifted gamma distribution with shape parameter $k > 0$, scale parameter $\lambda > 0$, and location parameter $\mu \in \mathbb{R} \setminus \{0\}$. We have

$$\frac{p'(x)}{p(x)} = \frac{k - 1}{x - \mu} - \frac{1}{\lambda}, \quad x > \mu.$$

Since $\mu \neq 0$, in order to establish our characterization result, we have to verify the conditions from Theorem 3 which includes (C3). However, for $k < 1$, we have

$$\int_\mu^\infty \big|p'(x)\big|\,\mathrm{d}x \ge \int_\mu^\infty \frac{|k - 1|}{x - \mu}\,p(x)\,\mathrm{d}x - \frac{1}{\lambda}\int_\mu^\infty p(x)\,\mathrm{d}x = \infty.$$

Next, we discuss the characterizations for probability distributions supported by the positive axis in the case of exponential families. More specifically, we focus on continuously differentiable density functions. Quite a few of the examples we already gave can be written as an exponential family, but we do not reconsider them and instead give a new example at the end of this part. Of course, the arguments below could also be used to treat exponential families over the real line, using Theorem 1. In detail, we let $\Theta \subset \mathbb{R}^d$ be non-empty and consider an exponential family (over the positive axis) in the natural parametrization given through

$$p_\vartheta(x) = c(\vartheta)\, h(x) \exp\left(\vartheta^\top T(x)\right), \quad x > 0,\ \vartheta \in \Theta,$$

where $T = (T_1, \ldots, T_d)^\top : (0, \infty) \to \mathbb{R}^d$ and $h : (0, \infty) \to [0, \infty)$ are (Borel) measurable functions, $\vartheta^\top$ is the transpose of a column vector ϑ, and

$$c(\vartheta) = \left(\int_0^\infty h(x) \exp\left(\vartheta^\top T(x)\right) \mathrm{d}x\right)^{-1}.$$

We choose Θ such that $0 < c(\vartheta) < \infty$ for each $\vartheta \in \Theta$. The exponential family is assumed to be strictly d-parametric, that is, we take the functions $1, T_1, \ldots, T_d$ to be linearly independent on the complement of every null set. The definition of exponential families and insights on their properties are provided by virtually any classical textbook on mathematical statistics.

We try to get an idea on how the conditions (C1)–(C4) can be handled for exponential families. Condition (C2) remains a little cryptic, meaning that it depends on the given example how it can be proven, and, at this point, we cannot give any improvement to what we discussed in Remark 7 concerning that condition.

(C1) Assume that T and h are continuously differentiable, and that h is positive. Trivially, these assumptions cover (C1) for they assure that for each $\vartheta \in \Theta$, p_ϑ is continuously differentiable and positive on $(0, \infty)$. For $x > 0$, we have

$$\frac{p_\vartheta'(x)}{p_\vartheta(x)} = \vartheta^\top T'(x) + \frac{h'(x)}{h(x)},$$

where $T'(x) = \left(T_1'(x), \ldots, T_d'(x)\right)^\top$.

(C3) Using the weaker subsidy for (C3) given in Remarks 5 and 6, a sufficient condition for (C3) is derived as follows. Let $\vartheta \in \Theta$, and take $Z \sim p_\vartheta \mathcal{L}^1$. Then,

$$\int_0^\infty x \left|p_\vartheta'(x)\right| \mathrm{d}x \le \|\vartheta\|\, \mathbb{E}\Big[\|T'(Z)\|\, Z\Big] + \mathbb{E}\left[\left|\frac{h'(Z)}{h(Z)}\right| Z\right].$$

Therefore, it suffices to know that

$$\mathbb{E}\left|\frac{h'(Z)}{h(Z)}\, Z\right| < \infty \quad \text{and} \quad \mathbb{E}\Big[\left|T_j'(Z)\right| Z\Big] < \infty, \quad j = 1, \ldots, d. \tag{7}$$

Since T often consists of monomials x^k, $k \in \mathbb{Z}$, or of some logarithmic term $\log(x)$, (7) frequently reduces to a moment constraint which is satisfied if the expectation of $T(Z)$ exists.

(C4) Note that P_ϑ, the distribution function corresponding to p_ϑ, trivially satisfies $\lim_{x \searrow 0} P_\vartheta(x) = 0$, so if $\lim_{x \searrow 0} p_\vartheta(x) > 0$, (C4) is obviously satisfied. If $\lim_{x \searrow 0} p_\vartheta(x) = 0$, a sufficient condition for (C4) is that

$$\lim_{x \searrow 0} \left(\vartheta^\top T'(x) + \frac{h'(x)}{h(x)} \right) = \infty.$$

We now give the characterization result that follows from Corollary 3. Corollary 2 yields a similar result via the density function, but we will not restate it explicitly.

Corollary 4 *Let $\{p_\vartheta \mid \vartheta \in \Theta\}$ be an exponential family as above. Assume that each p_ϑ is continuously differentiable and positive, and satisfies (C2)–(C4). Let X be a positive random variable with*

$$\mathbb{E}\left[\left(\left\| T'(X) \right\| + \left| \frac{h'(X)}{h(X)} \right| \right) X \right] < \infty.$$

Then, $X \sim p_\vartheta \mathcal{L}^1$ if, and only if, the distribution function of X has the form

$$F_X(t) = \mathbb{E}\left[-\left(\vartheta^\top T'(X) + \frac{h'(X)}{h(X)} \right) \min\{X, t\} \right], \quad t > 0.$$

Example 11 (*Log-normal distribution*) For parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, consider the density function of the log-normal distribution

$$\begin{aligned} p(x) &= \frac{1}{x\sqrt{2\pi}\,\sigma} \exp\left(-\frac{\left(\log(x) - \mu\right)^2}{2\sigma^2} \right) \\ &= \sqrt{-2\vartheta_2} \exp\left(\frac{\vartheta_1^2}{4\vartheta_2} \right) \frac{1}{\sqrt{2\pi}\, x} \exp\left(\vartheta_1 \log(x) + \vartheta_2 \log^2(x) \right), \quad x > 0, \end{aligned}$$

where $\vartheta = (\vartheta_1, \vartheta_2)^\top = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \right)^\top$. In the last representation, we see that the class of log-normal distributions forms an exponential family with parameter space $\Theta = \mathbb{R} \times (-\infty, 0)$, $h(x) = \frac{1}{\sqrt{2\pi}\, x}$, $T(x) = \left(\log(x), \log^2(x) \right)^\top$, as well as $c(\vartheta) = \sqrt{-2\vartheta_2} \exp\left(\frac{\vartheta_1^2}{4\vartheta_2} \right)$, where $c(\vartheta) \in (0, \infty)$ for every $\vartheta \in \Theta$. In this whole example, we suppress the index ϑ for p. All of the following arguments are valid for any fixed (but arbitrary) $\vartheta \in \Theta$.

The density function p is continuously differentiable since h and T are such, and it is positive as h is so. For $x > 0$, we have

$$\frac{p'(x)}{p(x)} = \vartheta^\top T'(x) + \frac{h'(x)}{h(x)} = \frac{\left(\mu - \sigma^2\right) - \log(x)}{\sigma^2 x}.$$

For the log-normal density function, we have $\lim_{x \searrow 0} p(x) = 0$, as well as

$$\lim_{x \searrow 0} \left(\vartheta^\top T'(x) + \frac{h'(x)}{h(x)} \right) = \lim_{x \searrow 0} \frac{\left(\mu - \sigma^2\right) - \log(x)}{\sigma^2 x} = \infty,$$

and the discussion of (C4) yields that this condition holds. In order to establish (C3), let $Z \sim p\mathcal{L}^1$. Then, $\log(Z)$ is Gaussian with mean μ and variance σ^2, and the expectation of $\log(Z)$ exists, that is, $\mathbb{E}|\log(Z)| < \infty$. Therefore, we have

$$\mathbb{E}\left|\frac{h'(Z)}{h(Z)}\,Z\right| = 1 < \infty, \quad \mathbb{E}\big|T_1'(Z)\,Z\big| = 1 < \infty,$$

and $\mathbb{E}\big|T_2'(Z)\,Z\big| = 2\,\mathbb{E}\big|\log(Z)\big| < \infty$, which suffices for (C3) by the discussions above. The proof of (C2) is a bit tedious and follows Remark 7. As it provides no insight on that regularity condition, we omit it here. The characterization result for the log-normal distribution as given in Corollary 4 is as follows. If X is a positive random variable with $\mathbb{E}|\log(X)| < \infty$, then X follows the log-normal law with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ if, and only if, the distribution function of X has the form

$$F_X(t) = \mathbb{E}\left[-\frac{(\mu - \sigma^2) - \log(X)}{\sigma^2 X}\,\min\{X, t\}\right], \quad t > 0.$$

Finally, we state the characterization result for a probability density function p with $\operatorname{spt}(p) = (-\infty, R]$, $R < \infty$. We omit the proof since it is a collage of earlier proofs, and only state the result. A characterization via the density function also holds, and its form is immediately conceivable from the result we state below. Similar observations concerning integrability conditions carry over from the case of density function with support bounded from below.

Corollary 5 *Let p be a probability density function with $\operatorname{spt}(p) = (-\infty, R]$ that satisfies (C1)–(C3) and (C5). Take $X : \Omega \to (-\infty, R)$ to be a random variable with $\mathbb{P}\big(X \in S(p)\big) = 1$, $\mathbb{E}\left|\frac{p'(X)}{p(X)}\right| < \infty$, and $\mathbb{E}\left|\frac{p'(X)}{p(X)}\,X\right| < \infty$. Then, $X \sim p\mathcal{L}^1$ if, and only if,*

$$1 - F_X(t) = \mathbb{E}\left[\frac{p'(X)}{p(X)}\Big(R - \max\{X, t\}\Big)\right], \quad t < R.$$

6 Univariate distributions with bounded support

For the sake of completeness, we study density functions $p : \mathbb{R} \to [0, \infty)$ with $\operatorname{spt}(p) = [L, R]$, where $L > -\infty$ and $R < \infty$. The proofs of our previous characterizations rely on the fact that $\lim_{x \to \pm\infty} p(x) = 0$. However, we can do more: The results can be extended to cases where the limit to one endpoint of the support merely exists. The techniques needed for the proofs of the statements in this section resemble the ones we have used so far, so we shorten the arguments. As in Sect. 5, we start with the characterizations via the density function before deriving further results from them. We divide the study into density functions for which the limit to the right endpoint of the support exists and such density functions for which the limit to the left endpoint exists.

Lemma 2 *Let p be a probability density function with $\operatorname{spt}(p) = [L, R]$ that satisfies (C1)–(C5), and for which the limit $\lim_{x \nearrow R} p(x)$ exists. Take a random variable*

$X : \Omega \to (L, R)$ with density function $\mathfrak{f}_X$, and $\mathbb{E}\left|\frac{p'(X)}{p(X)}\right| < \infty$. Then, $X \sim p\mathcal{L}^1$ if, and only if,

$$\mathfrak{f}_X(t) = \mathbb{E}\left[-\frac{p'(X)}{p(X)}\,\mathbb{1}\{X > t\}\right] + \lim_{x \nearrow R} p(x), \quad L < t < R.$$

The main ideas of the proof are summarized in the supplementary material, Section 4.

Remark 8 Note that condition (C3) is simply $\int_{S(p)} |p'(x)|\,\mathrm{d}x < \infty$ by the boundedness of the support. Also, notice that

$$\mathbb{E}\left|\frac{p'(X)}{p(X)}\,X\right| \le \max\left\{|L|, |R|\right\}\mathbb{E}\left|\frac{p'(X)}{p(X)}\right|,$$

so we never have to state both integrability conditions on X.

Remark 9 By the argument of Remark 5, in the case of a continuously differentiable density function with $L = 0$, we can replace the integrability condition on X completely with $\mathbb{E}\left|\frac{p'(X)}{p(X)}\,X\right| < \infty$ and substitute (C3) with $\int_0^R x|p'(x)|\,\mathrm{d}x < \infty$, which is weaker than $\int_0^R |p'(x)|\,\mathrm{d}x < \infty$. However, the equality

$$\mathfrak{f}_X(t) = \mathbb{E}\left[-\frac{p'(X)}{p(X)}\,\mathbb{1}\{X > t\}\right] + \lim_{x \nearrow R} p(x)$$

in Lemma 2 will then only hold for $\mathcal{L}^1$-a.e. $0 < t < R$.

Complementary to Lemma 2 (and with a similar proof), we have the following result.

Lemma 3 *Let p be a probability density function with $\mathrm{spt}(p) = [L, R]$ that satisfies (C1)–(C5), and for which the limit $\lim_{x \searrow L} p(x)$ exists. Assume that $X : \Omega \to (L, R)$ is a random variable with density function $\mathfrak{f}_X$, and $\mathbb{E}\left|\frac{p'(X)}{p(X)}\right| < \infty$. Then, $X \sim p\mathcal{L}^1$ if, and only if,*

$$\mathfrak{f}_X(t) = \mathbb{E}\left[\frac{p'(X)}{p(X)}\,\mathbb{1}\{X \le t\}\right] + \lim_{x \searrow L} p(x), \quad L < t < R.$$

With obvious adaptations, Remark 9 also applies here (in the case $R = 0$). We now use Lemmata 2 and 3 to derive the corresponding characterization results via the distribution function. We start again with the case of an existing limit to the right endpoint of the support.

Corollary 6 *Let p be a probability density function with $\mathrm{spt}(p) = [L, R]$ that satisfies (C1)–(C5). Assume that the limit $\lim_{x \nearrow R} p(x)$ exists. Take a random variable $X : \Omega \to (L, R)$ with $\mathbb{P}\big(X \in S(p)\big) = 1$, and $\mathbb{E}\left|\frac{p'(X)}{p(X)}\right| < \infty$. Then, $X \sim p\mathcal{L}^1$ if, and only if,*

$$F_X(t) = \mathbb{E}\left[-\frac{p'(X)}{p(X)}\Big(\min\{X, t\} - L\Big)\right] + (t - L)\lim_{x \nearrow R} p(x), \quad L < t < R.$$

The proof runs along the lines of Theorem 3.

Remark 10 Whenever p is continuously differentiable, and $L = 0$, it suffices to have $\int_0^R x|p'(x)|\,\mathrm{d}x < \infty$, instead of (C3), and the weaker condition $\mathbb{E}\left|\frac{p'(X)}{p(X)}\,X\right| < \infty$ to cover the requirements of Corollary 6.

The following result is complementary to Corollary 6.

Corollary 7 *Assume that p is a probability density function supported by $[L, R]$ that satisfies (C1)–(C5). Further suppose that $\lim_{x\searrow L} p(x)$ exists. Let $X : \Omega \to (L, R)$ be a random variable with $\mathbb{P}\big(X \in S(p)\big) = 1$, and $\mathbb{E}\left|\frac{p'(X)}{p(X)}\right| < \infty$. Then, $X \sim p\mathcal{L}^1$ if, and only if, the distribution function of X satisfies*

$$1 - F_X(t) = \mathbb{E}\left[\frac{p'(X)}{p(X)}\Big(R - \max\{X, t\}\Big)\right] + (R - t)\lim_{x\searrow L} p(x), \quad L < t < R.$$

Remark 10 applies, with minor adaptations, in the case $R = 0$. In general, the characterization results for probability density functions with bounded support give a good handle on a variety of wrapped and truncated distributions, like the truncated normal- or the wrapped exponential distribution. However, we state only the uniform- and the beta distribution as examples explicitly. Again, we refrain from giving the details of the calculations to check the regularity conditions. For the beta distribution, we invoke Remark 10.

Example 12 (*Uniform distribution*) For $x \in (L, R)$, let $p(x) = \frac{1}{R-L}$ be the density function of the uniform distribution on the interval (L, R). The conditions (C1)–(C5) are trivial to check. Since the derivate of p vanishes on (L, R), the identities from Corollaries 6 and 7 are the same. They read as follows. A random variable $X : \Omega \to (L, R)$ is distributed uniformly over (L, R) if, and only if, its distribution function has the form

$$F_X(t) = \frac{t - L}{R - L}, \quad L < x < R.$$

Apparently, we recovered the observation that the explicitly calculable form of the uniform distribution function uniquely identifies this distribution, so our characterization is redundant in this case.

Example 13 (*Beta distribution*) Let $\alpha > 0$, $\beta > 1$, and

$$p(x) = \frac{x^{\alpha-1}\,(1 - x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1,$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ denotes the beta function. Since $\beta > 1$, the limit to the right endpoint of the support exists. More precisely, we have that $\lim_{x\nearrow 1} p(x) = 0$. Therefore, Corollary 6 yields the following characterization. Suppose X is a random variable which takes values in $(0, 1)$ almost surely and satisfies $\mathbb{E}\left|\frac{X}{1-X}\right| < \infty$. Then,

X has the beta distribution with parameters $\alpha > 0$ and $\beta > 1$ if, and only if, the distribution function of X has the form

$$F_X(t) = \mathbb{E}\left[\left(\frac{\beta - 1}{1 - X} - \frac{\alpha - 1}{X}\right)\min\{X, t\}\right], \quad 0 < t < 1.$$

The beta distribution also marks a limitation of our characterizations. Namely, if $0 < \alpha, \beta < 1$, our results fail to hold since none of the required limits exist. A special case for this phenomenon is the arcsine distribution, which is the beta distribution with parameters $\alpha, \beta = \frac{1}{2}$.

7 Applications to goodness-of-fit testing

The idea to use distributional characterizations as a basis for statistics in testing problems is classic, see Nikitin (2017) and O'Reilly and Stephens (1982). In this spirit and regarding the results of the previous sections, we propose goodness-of-fit tests for any distribution with a density function that satisfies the regularity conditions of either of our characterizations (Theorems 1, 3, and Corollaries 5, 6, 7). For the sake of readability, we give the following discussion in the case of continuously differentiable and positive density functions on the positive axis dealt with in Corollary 3. This case includes the largest class of examples we gave previously, and it also includes the new test we provide. The arguments for using the characterizations for density functions on the whole real line or such densities that have bounded support to construct corresponding tests are very similar, of course.

We consider a parametric family of distributions $\mathcal{P} = \left\{p_\vartheta \mathcal{L}^1 \mid \vartheta \in \Theta\right\}$, $\Theta \subset \mathbb{R}^d$, where we assume that $\mathrm{spt}(p_\vartheta) = [0, \infty)$ and that p_ϑ is continuously differentiable and positive on $(0, \infty)$. Moreover, p_ϑ is taken to satisfy the prerequisites of Corollary 3. Testing the fit of a positive random variable X to $\mathcal{P}$ means to test the hypothesis

$$\mathbf{H_0} : \mathbb{P}^X \in \mathcal{P} \tag{8}$$

against general alternatives. Let $s : (0, \infty) \times \Theta \to (0, \infty)$ be a measurable function, used for scaling, such that $X \sim p_\vartheta \mathcal{L}^1$ if, and only if, $s(X; \vartheta) \sim p_{\vartheta^*} \mathcal{L}^1$ for some $\vartheta^* \in \Theta^* \subset \Theta$. We assume that

$$\mathbb{E}\left|\frac{p'_{\vartheta^*}\big(s(X;\vartheta)\big)}{p_{\vartheta^*}\big(s(X;\vartheta)\big)}\, s(X;\vartheta)\right| < \infty.$$

By Corollary 3, we have $s(X; \vartheta) \sim p_{\vartheta^*} \mathcal{L}^1$ if, and only if, the distribution function of $s(X; \vartheta)$ has the form

$$F_{s(X;\vartheta)}(t) = \mathbb{E}\left[-\frac{p'_{\vartheta^*}\big(s(X;\vartheta)\big)}{p_{\vartheta^*}\big(s(X;\vartheta)\big)}\, \min\left\{s(X;\vartheta),\, t\right\}\right], \quad t > 0. \tag{9}$$

In order to test $\mathbf{H_0}$ based on a sample $X_1, \ldots, X_n$ of independent and identically distributed (iid.) positive random variables, put $Y_{n,j} = s(X_j; \widehat{\vartheta}_n)$, for each $j = 1, \ldots, n$.

We consider the empirical distribution function $\widehat{F}_n$ of $Y_{n,1}, \ldots, Y_{n,n}$ as an estimator of $F_{s(X;\vartheta)}$. Hereby denoting a consistent estimator of ϑ by $\widehat{\vartheta}_n = \widehat{\vartheta}_n(X_1, \ldots, X_n)$, we use $\widehat{\vartheta}_n^* = \widehat{\vartheta}_n(Y_{n,1}, \ldots, Y_{n,n})$ as an estimator of $\vartheta^* \in \Theta^*$ and take

$$\widehat{T}_n(t) = -\frac{1}{n} \sum_{j=1}^{n} \frac{p'_{\widehat{\vartheta}_n^*}\big(Y_{n,j}\big)}{p_{\widehat{\vartheta}_n^*}\big(Y_{n,j}\big)} \min\big\{Y_{n,j}, t\big\}, \quad t > 0,$$

as an estimator of the second quantity in (9). Taking some metric δ on a set containing both functions, we propose as a goodness-of-fit statistic the quantity $\delta\big(\widehat{T}_n, \widehat{F}_n\big)$. By (9), this term ought to be close to zero under $\mathbf{H_0}$, so large values of the statistic will lead us to rejecting the hypothesis.

As witnessed by Baringhaus and Henze (2000), Betsch and Ebner (2019a, b), tests of this type are noteworthy competitors to established tests. An advantage lies in the range of their applicability. A substantial proportion of known procedures relies on a comparison between theoretical moment generating functions, see Cabaña and Quiroz (2005), Henze and Jiménez-Gamero (2019) and Zghoul (2010), or characteristic functions, see Baringhaus and Henze (1988), Epps and Pulley (1983) and Jiménez-Gamero et al. (2009), and their empirical pendants, or on a differential equation that characterizes the Laplace transformation, see Henze and Klar (2002) and Henze et al. (2012). All of these share the unpleasant feature that in order to establish the theoretic basis for the test statistics, one has to have explicit knowledge about these transformations for the distribution in consideration. Since their handling is not possible for every distribution, our suggestions provide a genuine alternative, for they require no more than the knowledge of the density function and its derivative. Moreover, our tests do not rely on a characterization that is tailored to one specific distribution. Instead, we provide a framework for testing fit to many different distributions. To illustrate the technical setting above, we add to the supplementary material (Sections 5 and 6) insight into how the tests for the gamma and normal distribution by Betsch and Ebner (2019a, b), respectively, fit into this framework. The test for exponentiality by Baringhaus and Henze (2000) also appears as a special case. Also, we provide further references which show that these testing problems are still of interest to researchers. In Section 7 of the supplementary material, we also explain that the Kolmogorov–Smirnov and Cramér-von Mises tests fit into our framework.

In the remainder of this work, we propose a new goodness-of-fit test for the two-parameter Burr Type XII distribution $\mathrm{Burr}_{\mathrm{XII}}(k, c)$, $k, c > 0$, based on the characterization given in Example 7, fixing the scale parameter $\sigma = 1$. The distribution is known under a variety of names, e.g., as the Singh–Maddala distribution or as the Pareto IV distribution, for details see Kleiber and Kotz (2003), Section 6.2. We denote the density function of the $\mathrm{Burr}_{\mathrm{XII}}(k, c)$ distribution by $p_\vartheta(x) = c\, k\, x^{c-1}(1+x^c)^{-k-1}$, $x > 0$, with parameter vector $\vartheta = (k, c) \in (0, \infty)^2 = \Theta$. For iid. copies $X_1, \ldots, X_n$ of X, define

$$\widehat{T}_n(t) = \frac{1}{n} \sum_{j=1}^{n} \left(\widehat{c}_n \left(\widehat{k}_n + 1\right) \frac{X_j^{\widehat{c}_n - 1}}{1 + X_j^{\widehat{c}_n}} - \frac{\widehat{c}_n - 1}{X_j} \right) \min\{X_j, t\}, \quad t > 0,$$

in accordance with the general framework above, which leads to the family of L^2-type statistics

$$B_{n,a} = n \int_0^\infty \left|\widehat{T}_n(t) - \widehat{F}_n(t)\right|^2 w_a(t)\,\mathrm{d}t.$$

Here, $\widehat{k}_n$ and $\widehat{c}_n$ are consistent estimators of the parameters k and c, $\widehat{F}_n$ is the empirical distribution function of $X_1, \ldots, X_n$, and $w_a(t) = \exp(-at)$, $t > 0$, is a weight function depending on a tuning parameter $a > 0$. Rejection of the hypothesis $\mathbf{H_0}$ in (8), i.e., that the data come from the Burr Type XII family, is for large values of $B_{n,a}$. Writing $X_{(1)} \leq \ldots \leq X_{(n)}$ for the order statistics of $X_1, \ldots, X_n$, we have after some tedious calculations

$$\begin{aligned}
B_{n,a} = \frac{2}{n} \sum_{1\leq j<\ell\leq n} \Bigg\{ A^{[1]}_{(\ell),n} \Bigg[& \frac{2A^{[1]}_{(j),n}}{a^3}\left(1 - e^{-aX_{(j)}}\right) + \frac{A^{[2]}_{(j),n}}{a^2}\left(e^{-aX_{(j)}} + e^{-aX_{(\ell)}}\right) \\
& + \frac{\widehat{c}_n - 2}{a^2}\, e^{-aX_{(j)}} - \frac{X_{(j)}}{a}\, e^{-aX_{(j)}} \Bigg] + \frac{A^{[2]}_{(j),n}}{a}\, e^{-aX_{(\ell)}} \Bigg\} \\
+ \frac{1}{n}\sum_{j=1}^n \Bigg\{ \left(A^{[1]}_{(j),n}\right)^2 & \left(-\frac{2X_{(j)}}{a^2}\, e^{-aX_{(j)}} - \frac{2}{a^3}\, e^{-aX_{(j)}} + \frac{2}{a^3} \right) \\
& + \frac{2(j-1)\,\widehat{c}_n}{a^2}\, A^{[1]}_{(j),n}\, e^{-aX_{(j)}} + \frac{2A^{[2]}_{(j),n}}{a}\, e^{-aX_{(j)}} \Bigg\} \\
+ \frac{2\widehat{c}_n}{a\,n} \sum_{j=1}^n j\, e^{-aX_{(j)}} & - \frac{1}{a\,n}\sum_{j=1}^n e^{-aX_{(j)}},
\end{aligned}$$

where

$$A^{[1]}_{(j),n} = \widehat{c}_n\,\left(\widehat{k}_n + 1\right)\,\frac{X_{(j)}^{\widehat{c}_n - 1}}{1 + X_{(j)}^{\widehat{c}_n}} - \frac{\widehat{c}_n - 1}{X_{(j)}}, \quad A^{[2]}_{(j),n} = -\widehat{c}_n\,\left(\widehat{k}_n + 1\right)\,\frac{X_{(j)}^{\widehat{c}_n}}{1 + X_{(j)}^{\widehat{c}_n}},$$

which is an easily computable formula that avoids any numerical integration routines. In the following simulation study, we show the effectiveness of this new test statistics in comparison with the classical procedures adapted for the composite hypothesis $\mathbf{H_0}$, namely the Kolmogorov–Smirnov test K_n, the Cramér-von Mises test CM, the Anderson–Darling test AD and the Watson test WA. Let $F(x; k, c) = 1 - (1 + x^c)^{-k}$, $x > 0$, denote the distribution function of $\mathrm{Burr_{XII}}(k, c)$. The K_n-statistic is equal to $K_n = \max\{D^+, D^-\}$, where

$$\begin{aligned}
D^+ &= \max_{j=1,\ldots,n}\left(j/n - F\left(X_{(j)}; \widehat{k}_n, \widehat{c}_n\right)\right), \\
D^- &= \max_{j=1,\ldots,n}\left(F\left(X_{(j)}; \widehat{k}_n, \widehat{c}_n\right) - (j-1)/n\right).
\end{aligned}$$

The statistics of Cramér-von Mises and Anderson–Darling are given by

$$\mathrm{CM} = \frac{1}{12n} + \sum_{j=1}^{n} \left(F\left(X_{(j)}; \widehat{k}_n, \widehat{c}_n\right) - \frac{2j-1}{2n} \right)^2$$

and

$$\mathrm{AD} = -n - \frac{1}{n} \sum_{j=1}^{n} \Big[(2j-1) \log F\Big(X_{(j)}; \widehat{k}_n, \widehat{c}_n\Big) + \big(2(n-j)+1\big) \log\Big(1 - F\Big(X_{(j)}; \widehat{k}_n, \widehat{c}_n\Big)\Big) \Big],$$

respectively, whereas the WA-statistic takes the form

$$\mathrm{WA} = \mathrm{CM} - n \left(\frac{1}{n} \sum_{j=1}^{n} F\left(X_{(j)}; \widehat{k}_n, \widehat{c}_n\right) - \frac{1}{2} \right)^2.$$

For all procedures, the parameters are estimated via the maximum likelihood method, maximizing numerically the log-likelihood function, see Jalali and Watkins (2009) and Wingo (1983). There are other estimation procedures available, like the maximum product of spacings method, see Shah and Gokhale (1993). Critical points are obtained for the classical tests, as well as for the new test, by the same parametric bootstrap procedure, as follows: For a given sample $X_1, \ldots, X_n$ of size n, compute the estimators $\widehat{k}_n, \widehat{c}_n$ of k and c. Conditionally on $\widehat{k}_n, \widehat{c}_n$, generate 100 bootstrap samples of size n from $\mathrm{Burr}_{\mathrm{XII}}(\widehat{k}_n, \widehat{c}_n)$. Calculate the value of the test statistic, say B_j^*, $(j = 1, \ldots, 100)$, for each bootstrap sample. Obtain the critical value p_n as $B_{(90)}^*$, where $B_{(j)}^*$ denote the ordered B_j^*-values, and reject the hypothesis $\mathbf{H_0}$ if $B_{n,a} = B_{n,a}(X_1, \ldots, X_n) > p_n$.

The following (alternative) distributions are considered (all densities defined for $x > 0$ in dependence of a shape parameter $\theta > 0$):

1. The Burr Type XII distribution $\mathrm{Burr}_{\mathrm{XII}}(k, c)$,
2. the exponential distribution $\mathrm{Exp}(\theta)$,
3. the linear increasing failure rate law $LF(\theta)$ with density function given as $(1 + \theta x)\exp(-x - \theta x^2/2)$,
4. the half-normal distribution with density function $(2/\pi)^{1/2} \exp(-x^2/2)$, denoted by HN,
5. the half-Cauchy distribution with density $2/\big(\pi(1 + x^2)\big)$, denoted by HC,
6. the Gompertz law $GO(\theta)$ having distribution function $1 - \exp[\theta^{-1}(1 - e^x)]$,
7. the inverse Gaussian distribution $IG(\theta)$ with density function given through $\big(\theta/(2\pi)\big)^{1/2} x^{-3/2} \exp[-\theta(x-1)^2/(2x)]$,
8. the Weibull distribution with density $\theta x^{\theta-1} \exp(-x^\theta)$, denoted by $W(\theta)$,
9. the inverse Weibull distribution with density $\theta(1/x)^{\theta+1} \exp[-(1/x)^\theta]$, denoted by $IW(\theta)$.

Table 1 Percentage of rejection for 10,000 Monte Carlo repetitions ($n = 100$, $\alpha = 0.1$)

Alt./Test	$B_{0.25}$	$B_{0.5}$	B_1	B_3	B_5	B_{10}	K_n	CM	AD	WA
$\mathrm{Burr}_{\mathrm{XII}}(1, 1)$	10	10	10	10	10	10	10	10	10	11
$\mathrm{Burr}_{\mathrm{XII}}(2, 1)$	9	10	10	10	10	10	10	10	09	11
$\mathrm{Burr}_{\mathrm{XII}}(4, 1)$	6	7	8	11	10	10	11	10	10	11
$\mathrm{Burr}_{\mathrm{XII}}(0.5, 2)$	9	9	9	10	9	8	10	10	10	10
$\mathrm{Burr}_{\mathrm{XII}}(2, 0.5)$	10	10	10	10	10	11	10	10	10	10
Exp(1)	0	27	**69**	55	44	42	51	61	66	53
LF(2)	0	0	2	**77**	73	63	56	67	74	59
LF(4)	0	0	0	56	**68**	60	48	57	64	46
HC	12	12	13	14	13	12	12	13	**15**	14
HN	0	1	64	89	81	73	78	88	**90**	78
GO(2)	0	4	90	99	98	93	97	99	**100**	98
IG(0.5)	13	45	66	81	**83**	82	52	64	72	61
IG(1.5)	2	6	22	40	**48**	46	24	31	37	32
IG(3)	1	2	8	18	**24**	24	16	21	23	23
$W(0.5)$	**74**	70	60	32	23	22	52	61	65	53
$W(1.5)$	0	0	38	**66**	54	47	52	60	65	52
$W(3)$	0	0	0	65	**69**	56	52	61	65	52
IW(1)	46	50	56	**66**	63	44	37	42	48	41

All computations are performed using the statistical computing environment R, see R Core Team (2019). In each scenario, we consider the sample sizes $n = 100$ and $n = 200$, and the nominal level of significance α is set to 0.1. Each entry in Tables 1 and 2 presents empirical rejection rates computed with 10,000 Monte Carlo runs. The number of bootstrap samples in each run is fixed to 100, and for the tuning parameter a, we consider the values $\{0.25, 0.5, 1, 3, 5, 10\}$. The best performing test for each distribution and sample size is highlighted for easy reference.

The simulation results show the (strong) dependence on the tuning parameter, but also, for an appropriate choice, the effectiveness of the new procedures, outperforming the classical procedures almost uniformly with the exception of the half-Cauchy-, half-normal- and Gompertz distribution, where the Anderson–Darling statistic is the most powerful test. Clearly, a data-dependent choice for an optimal tuning parameter is desirable. Unfortunately, there is no known procedure for this kind of test statistics, where the distribution under $\mathbf{H_0}$ depends on the true values of the parameters, but results for tests of location–scale families by Allison and Santana (2015) and Tenreiro (2019) give hope for new developments. A compromise for practitioners concerning the choice of the tuning parameter is $a = 3$ in view of Tables 1 and 2.

Similar to the other procedures based on our approach, see Sections 5 and 6 in the supplementary material, we expect the statistics $B_{n,a}$ to converge under $\mathbf{H_0}$ to the L^2-norm of a centered Gaussian process and the tests to be consistent against fixed alternatives.

Table 2 Percentage of rejection for 10,000 Monte Carlo repetitions ($n = 200$, $\alpha = 0.1$)

Alt./Test	$B_{0.25}$	$B_{0.5}$	B_1	B_3	B_5	B_{10}	K_n	CM	AD	WA
$\text{Burr}_{\text{XII}}(1, 1)$	10	10	10	10	10	10	10	10	10	10
$\text{Burr}_{\text{XII}}(2, 1)$	10	10	10	10	10	11	10	9	9	10
$\text{Burr}_{\text{XII}}(4, 1)$	7	8	10	10	10	10	10	10	10	11
$\text{Burr}_{\text{XII}}(0.5, 2)$	10	10	10	10	10	9	11	10	10	9
$\text{Burr}_{\text{XII}}(2, 0.5)$	10	10	9	10	9	10	10	10	10	10
Exp(1)	4	78	**95**	83	71	66	81	88	92	82
LF(2)	0	0	19	**97**	95	89	85	92	95	88
LF(4)	0	0	0	89	**91**	85	75	83	89	76
HC	14	13	15	**19**	18	15	15	17	**19**	18
HN	0	17	98	**100**	98	94	97	99	**100**	98
GO(2)	0	57	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
IG(0.5)	29	78	93	**99**	**99**	**99**	82	93	97	92
IG(1.5)	1	13	41	72	82	**85**	44	53	66	56
IG(3)	0	2	13	34	48	**61**	28	37	46	40
$W(0.5)$	**97**	94	86	55	40	36	81	89	91	80
$W(1.5)$	0	6	84	91	81	70	80	88	**92**	82
$W(3)$	0	0	6	93	**94**	88	80	89	92	82
IW(1)	74	78	83	**93**	**93**	85	63	72	78	69

8 Conclusions

We devoted this work to the derivation of explicit characterizations for a large class of continuous univariate probability distributions. Our motivation was the fact that the characterization of the standard normal distribution as the unique fixed point of the zero-bias transformation reduces to an explicit formula for the distribution function of the transformed distribution. We extrapolated this formula to other distributions by applying the Stein-type identity commonly used within the density approach. Research related to our characterizations concerns the study of distributional transformations, see Goldstein and Reinert (2005) and Döbler (2017). While these are constructed from scratch and are used to prove Stein-type characterizations, we took such a Stein identity for granted and dropped the ambition to obtain distributional transformations. Thus, starting with more information and demanding less structure from the transformations, we established better accessible explicit characterization formulae. In the last section, we discussed an immediate application. We illustrated how to use the characterizations for the construction of goodness-of-fit tests. The corresponding procedures for the normal-, the exponential- and the gamma distribution have already been investigated in the literature, and they show very promising performance. The great advantage of our approach lies in the wide range of its applicability. To confirm this last claim, we constructed the (to our best knowledge) first ever goodness-of-fit test focused on the Burr Type XII distribution.

Acknowledgements The authors would like to thank an associate editor as well as three anonymous reviewers for their comments and suggestions that led to a major improvement of the paper.

References

Allison, J. S., Santana, L. (2015). On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *Journal of Statistical Computation and Simulation*, *85*(16), 3276–3288.

Anastasiou, A. (2018). Assessing the multivariate normal approximation of the maximum likelihood estimator from high-dimensional, heterogeneous data. *Electronic Journal of Statistics*, *12*(2), 3794–3828.

Anastasiou, A., Gaunt, R. (2019). Multivariate normal approximation of the maximum likelihood estimator via the delta method. to appear in *Brazilian Journal of Probability and Statistics* arXiv:1609.03970.

Anastasiou, A., Reinert, G. (2017). Bounds for the normal approximation of the maximum likelihood estimator. *Bernoulli*, *23*(1), 191–218.

Anastasiou, A., Reinert, G. (2018). Bounds for the asymptotic distribution of the likelihood ratio. arXiv e-prints arXiv:1806.03666.

Barbour, A. D. (1982). Poisson convergence and random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society*, *92*(2), 349–359.

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

Barbour, A. D., Karoński, M., Ruciński, A. (1989). A central limit theorem for decomposable random variables with applications to random graphs. *Journal of Combinatorial Theory, Series B*, *47*(2), 125–145.

Baringhaus, L., Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, *35*(1), 339–348.

Baringhaus, L., Henze, N. (2000). Tests of fit for exponentiality based on a characterization via the mean residual life function. *Statistical Papers*, *41*(2), 225–236.

Betsch, S., Ebner, B. (2019a). A new characterization of the Gamma distribution and associated goodness-of-fit tests. *Metrika*, *82*(7), 779–806.

Betsch, S., Ebner, B. (2019b). Testing normality via a distributional fixed point property in the Stein characterization. *TEST*, https://doi.org/10.1007/s11749-019-00630-0.

Braverman, A., Dai, J. G. (2017). Stein's method for steady-state diffusion approximations of $M/Ph/n+M$ systems. *The Annals of Applied Probability*, *27*(1), 550–581.

Braverman, A., Dai, J. G., Feng, J. (2016). Stein's method for steady-state diffusion approximations: An introduction through the Erlang-A and Erlang-C models. *Stochastic Systems*, *6*(2), 301–366.

Cabaña, A., Quiroz, A. (2005). Using the empirical moment generating function in testing for the Weibull and the type I extreme value distributions. *TEST*, *14*(2), 417–432.

Carrillo, C., Cidrás, J., Díaz-Dorado, E., Obando-Montaño, A. F. (2014). An approach to determine the Weibull parameters for wind energy analysis: The case of Galicia (Spain). *Energies*, *7*(4), 2676–2700.

Chatterjee, S., Shao, Q.-M. (2011). Nonnormal approximation by Stein's method of exchangeable pairs with application to the Curie–Weiss model. *The Annals of Applied Probability*, *21*(2), 464–483.

Chen, L. H. Y., Goldstein, L., Shao, Q.-M. (2011). *Normal approximation by Stein's method*. Berlin: Springer.

Chwialkowski, K., Strathmann, H., Gretton, A. (2016). A kernel test of goodness of fit. *Proceedings of the 33rd international conference on machine learning, ICML'16* (Vol. 48, pp. 2606–2615).

Döbler, C. (2015). Stein's method of exchangeable pairs for the Beta distribution and generalizations. *Electronic Journal of Probability*, *20*(109), 1–34.

Döbler, C. (2017). Distributional transformations without orthogonality relations. *Journal of Theoretical Probability*, *30*(1), 85–116.

Epps, T. W., Pulley, L. B. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, *70*(3), 723–726.

Fang, X. (2014). Discretized normal approximation by Stein's method. *Bernoulli*, *20*(3), 1404–1431.

Gaunt, R., Pickett, A., Reinert, G. (2017). Chi-square approximation by Stein's method with application to Pearson's statistic. *Annals of Applied Probability*, *27*(2), 720–756.

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

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

Götze, F. (1991). On the rate of convergence in the multivariate CLT. *The Annals of Probability*, *19*(2), 724–739.

Henze, N., Jiménez-Gamero, M. D. (2019). A new class of tests for multinormality with iid and garch data based on the empirical moment generating function. *TEST*, *28*(2), 499–521.

Henze, N., Klar, B. (2002). Goodness-of-fit tests for the inverse Gaussian distribution based on the empirical Laplace transform. *Annals of the Institute of Statistical Mathematics*, *54*(2), 425–444.

Henze, N., Meintanis, S. G., Ebner, B. (2012). Goodness-of-fit tests for the Gamma distribution based on the empirical Laplace transform. *Communications in Statistics-Theory and Methods*, *41*(9), 1543–1556.

Hudson, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *The Annals of Statistics*, *6*(3), 473–484.

Jalali, A., Watkins, A. J. (2009). On maximum likelihood estimation for the two parameter Burr XII distribution. *Communications in Statistics—Theory and Methods*, *38*(11), 1916–1926.

Jiménez-Gamero, M. D., Alba-Fernández, V., Muñoz-García, J., Chalco-Cano, Y. (2009). Goodness-of-fit tests based on empirical characteristic functions. *Computational Statistics & Data Analysis*, *53*(12), 3957–3971.

Kim, S.-T. (2000). A use of the Stein-Chen method in time series analysis. *Journal of Applied Probability*, *37*(4), 1129–1136.

Kleiber, C., Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*. Wiley series in probability and statistics. Hoboken: Wiley.

Ley, C. and Swan, Y. (2011). A unified approach to Stein characterizations. arXiv e-prints arXiv:1105.4925v3.

Ley, C., Swan, Y. (2013a). Local Pinsker inequalities via Stein's discrete density approach. *IEEE Transactions on Information Theory*, *59*(9), 5584–5591.

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

Ley, C., Swan, Y. (2016). Parametric Stein operators and variance bounds. *Brazilian Journal of Probability and Statistics*, *30*(2), 171–195.

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

Linnik, Y. V. (1962). Linear forms and statistical criteria I, II. *Selected Translations in Mathematical Statistics and Probability, 3*,1–40: 41–90. Originally published 1953 in the Ukrainian Mathematical Journal, Vol. 5, pp. 207–243, 247–290 **(in Russian)**.

Liu, Q., Lee, J. D., Jordan, M. (2016). A kernelized Stein discrepancy for goodness-of-fit tests. *Proceedings of the 33rd International Conference on Machine Learning*, ICML'16, (Vol. 46, pp. 276–284).

Nikitin, Y. Y. (2017). Tests based on characterizations, and their efficiencies: A survey. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, *21*(1), 3–24.

O'Reilly, F. J., Stephens, M. A. (1982). Characterizations and goodness of fit tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, *44*(3), 353–360.

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

Pinelis, I. (2017). Optimal-order uniform and nonuniform bounds on the rate of convergence to normality for maximum likelihood estimators. *Electronic Journal of Statistics*, *11*(1), 1160–1179.

Prakasa Rao, B. L. S. (1979). Characterizations of distributions through some identities. *Journal of Applied Probability*, *16*(4), 903–909.

Proakis, J. G., Salehi, M. (2008). *Digital communications*, 5th ed. New York: McGraw-Hill.

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Reinert, G., Röllin, A. (2010). Random subgraph counts and U-statistics: Multivariate normal approximation via exchangeable pairs and embedding. *Journal of Applied Probability*, *47*(2), 378–393.

Rogers, G. L. (2008). Multiple path analysis of reflectance from turbid media. *Journal of the Optical Society of America A*, *25*(11), 2879–2883.

Ross, N. (2011). Fundamentals of Stein's method. *Probability Surveys*, *8*, 210–293.

Shah, A., Gokhale, D. V. (1993). On maximum product of spacings (mps) estimation for Burr XII distributions. *Communications in Statistics—Simulation and Computation*, *22*(3), 615–641.

Singh, S. K., Maddala, G. S. (1976). A function for size distribution of incomes. *Econometrica*, *44*(5), 963–970.
Singh, V. P. (1987). On application of the Weibull distribution in hydrology. *Water Resources Management*, *1*(1), 33–43.
Stein, C. (1986). *Approximate computation of expectations*, Vol. 7. Hayward: Institute of Mathematical Statistics.
Stein, C., Diaconis, P., Holmes, S., Reinert, G. (2004). Use of exchangeable pairs in the analysis of simulations. In P. Diaconis & S. Holmes (Eds.), *Stein's method. Lecture notes-monograph series*, Vol. 46, pp. 1–25. Beachwood, OH: Institute of Mathematical Statistics.
Tenreiro, C. (2019). On the automatic selection of the tuning parameter appearing in certain families of goodness-of-fit tests. *Journal of Statistical Computation and Simulation*, *89*(10), 1780–1797.
Wingo, D. R. (1983). Maximum likelihood methods for fitting the Burr type XII distribution to life test data. *Biometrical Journal*, *25*(1), 77–84.
Ying, L. (2017). Stein's method for mean-field approximations in light and heavy traffic regimes. *SIGMETRICS 2017 abstracts—Proceedings of the 2017 ACM SIGMETRICS/International conference on measurement and modeling of computer systems*. Association for Computing Machinery, Inc.
Zghoul, A. A. (2010). A goodness of fit test for normality based on the empirical moment generating function. *Communications in Statistics—Simulation and Computation*, *39*(6), 1292–1304.