Michael Falk* and Gilles Stupfler

On a class of norms generated by nonnegative integrable distributions

Abstract: We show that any distribution function on $\mathbb{R}^d$ with nonnegative, nonzero and integrable marginal distributions can be characterized by a norm on $\mathbb{R}^{d+1}$, called F-norm. We characterize the set of F-norms and prove that pointwise convergence of a sequence of F-norms to an F-norm is equivalent to convergence of the pertaining distribution functions in the Wasserstein metric. On the statistical side, an F-norm can easily be estimated by an empirical F-norm, whose consistency and weak convergence we establish.

The concept of F-norms can be extended to arbitrary random vectors under suitable integrability conditions fulfilled by, for instance, normal distributions. The set of F-norms is endowed with a semigroup operation which, in this context, corresponds to ordinary convolution of the underlying distributions. Limiting results such as the central limit theorem can then be formulated in terms of pointwise convergence of products of F-norms.

We conclude by showing how, using the geometry of F-norms, we may characterize nonnegative integrable distributions in $\mathbb{R}^d$ by simple compact sets in $\mathbb{R}^{d+1}$. We then relate convergence of those distributions in the Wasserstein metric to convergence of these characteristic sets with respect to Hausdorff distances.

Keywords: Characteristic function, D-norm, empirical distribution function, Hausdorff metric, multivariate distribution, norm, Wasserstein metric

MSC: 60E10, 60G99, 62H05, 62H12

1 Introduction

It was observed only recently that a particular kind of norms on $\mathbb{R}^d$, called *D-norms*, are the skeleton of multivariate extreme value theory. An up-to-date account of D-norms is [4]. D-norms are defined via a random vector (rv), called *generator*. The distribution function (df) of this rv, however, is not uniquely determined, and there exists an infinite number of generators of the same D-norm. It was shown by [6] that the D-norm characterizes the distribution of a generator if the constant function one is added to the generator as a further component. This led to the definition of the *max-characteristic function*, which can be used to identify the distribution of any multivariate distribution with nonnegative and integrable components.

In this paper we build on these observations and construct a norm on $\mathbb{R}^{d+1}$, called F-norm, which contains the notion of max-characteristic function. In Section 2.1, we present the concept of F-norms, and show that the df of each rv $\boldsymbol{X} = (X_1, \ldots, X_d)$ on $\mathbb{R}^d$ with nonnegative, nonzero and integrable components can be characterized by the pertaining F-norm. We then list examples and derive basic properties as well as an inversion formula to retrieve a distribution from its associated F-norm. We also fully characterize the set of F-norms and obtain a simple classification in two dimensions.

*Corresponding Author: Michael Falk:** Institute of Mathematics, University of Würzburg, Würzburg, Germany, E-mail:
michael.falk@uni-wuerzburg.de
Gilles Stupfler: School of Mathematical Sciences, University of Nottingham, Nottingham, UK, E-mail:
gilles.stupfler@gmail.com

In Section 3 we analyse the convergence of sequences of F-norms. We start by proving that pointwise convergence of a sequence of F-norms to an F-norm is equivalent with convergence of the pertaining dfs with respect to the Wasserstein metric. We then add some statistical views on F-norms to this section. The (random) F-norm $\|\cdot\|_{\widehat{F}_n}$ of the empirical df $\widehat{F}_n$ of a sample of n independent and identically distributed (iid) rvs is an estimator of $\|\cdot\|_F$ with the structure of a sample mean. Local uniform consistency and asymptotic normality of $\|\cdot\|_{\widehat{F}_n}$ as an estimator of $\|\cdot\|_F$ are then consequences of the law of large numbers and the multivariate central limit theorem. More strongly, we establish the $\sqrt{n}$-functional weak convergence of $\|\cdot\|_{\widehat{F}_n} - \|\cdot\|_F$ to a Gaussian process which is essentially a functional of a Brownian bridge.

Section 3 suggests that F-norms interact nicely with well-known modes of convergence and theorems of statistical analysis. In order to be able to use these norms in practice for asymptotic analyses, it is important to understand how they behave with respect to simple algebraic operations. It turns out that two F-norms can be multiplied by constructing the F-norm generated by the componentwise product of pairs of *independent* rvs giving rise to the individual F-norms. We also provide an integral formula making it possible, given two F-norms, to compute this product in a straightforward way. Equipped with this commutative multiplication, the set of F-norms is a semigroup with an identity element, and we can fully identify the invertible and idempotent elements for this operation. This algebraic aspect is investigated in Section 4.

The concept of F-norms as we introduce it originally focuses on multivariate rvs with nonnegative and integrable components, and thus excludes common distributions such as the multivariate normal distribution. In Section 5 we show that we can also define, by an exponential transformation, a concept of F-norms for a rv attaining negative values, under an integrability condition. This indeed allows us to include multivariate normal distributions, as well as other interesting examples. The multiplication of F-norms in Section 4 then represents the convolution of two rvs, and central limit theorems for iid rvs now mean pointwise convergence of the sequence of corresponding products of F-norms.

A multivariate distribution can then, under an integrability assumption, be characterized by its associated F-norm. The norm structure makes it possible to reduce the knowledge of the df F to even simpler objects than the full F-norm. Because each norm is a homogeneous function, the knowledge of an F-norm (and thus of the underlying df F) is equivalent to its knowledge on the unit simplex. Besides, and since a norm is characterized by its unit sphere, multivariate distributions on $\mathbb{R}^d$ can be characterized, under suitable integrability conditions on the components, by the part of the unit sphere for their F-norm contained in the positive orthant of $\mathbb{R}^{d+1}$, which is a compact set. Interestingly, the convergence of F-norms, and therefore convergence of d-dimensional distributions in the Wasserstein metric, can be shown to be equivalent to the convergence of these unit spheres with respect to any Hausdorff metric induced by a norm in $\mathbb{R}^{d+1}$. These geometric aspects are investigated in Section 6. Section 7 concludes.

2 The concept of F-norms

2.1 Definition, examples, and basic properties

Let $d \geq 1$ and $\boldsymbol{X} = (X_1, \dots, X_d)$ be a rv satisfying the fundamental assumption

($\mathcal{H}$) Each X_i is almost surely (a.s.) nonnegative with $0 < E(X_i) < \infty$.

Denote by F the df of $\boldsymbol{X}$. For $\boldsymbol{x} = (x_0, x_1, \dots, x_d) \in \mathbb{R}^{d+1}$, define a mapping $\|\cdot\|_F$ by

$$\begin{aligned}\|\boldsymbol{x}\|_F &:= E\left(\max\left(|x_0|, |x_1| X_1, \dots, |x_d| X_d\right)\right) \\ &= E\left(\left\|(x_0, x_1 X_1, \dots, x_d X_d)\right\|_\infty\right).\end{aligned} \tag{1}$$

As shown in Theorem 2.2 below, the distribution of $\boldsymbol{X}$ is characterized by the mapping $\|\cdot\|_F$. This is not true in general if we replace the sup-norm $\|\cdot\|_\infty$ in the above definition (1) by an arbitrary norm. Indeed, consider

for instance the L^1-norm and define

$$\|\boldsymbol{x}\|_F^{\star} := E\left(|x_0| + \sum_{i=1}^{d} |x_i| X_i\right) = |x_0| + \sum_{i=1}^{d} |x_i| E(X_i).$$

An *arbitrary* rv $\boldsymbol{X} \geq \boldsymbol{0} \in \mathbb{R}^d$ with $E(X_i) = 1$, $1 \leq i \leq d$, then provides the same value $\|\boldsymbol{x}\|_F^{\star}$ as the constant rv $\boldsymbol{X} = (1, \ldots, 1) \in \mathbb{R}^d$. The use of the sup-norm $\|\cdot\|_\infty$ in definition (1) is, therefore, crucial.

This paper is based on the following fundamental observations, presented in the two subsequent results. The proof of the first result is elementary.

Lemma 2.1. *If $\boldsymbol{X}$ satisfies ($\mathcal{H}$) then $\|\cdot\|_F$ is a norm on $\mathbb{R}^{d+1}$.*

Moreover, the norm $\|\cdot\|_F$ characterizes the df of $\boldsymbol{X}$, which can be seen as follows. The function $\varphi_{\boldsymbol{X}}$, defined for any $\boldsymbol{x} = (x_1, \ldots, x_d) \geq \boldsymbol{0} \in \mathbb{R}^d$ by

$$\varphi_{\boldsymbol{X}}(\boldsymbol{x}) := E\left(\max(1, x_1 X_1, \ldots, x_d X_d)\right), \qquad (2)$$

is the *max-characteristic function* (max-CF) pertaining to $\boldsymbol{X}$ (any operation on vectors such as $+, \geq, \ldots$ is meant componentwise throughout). It characterizes the distribution of $\boldsymbol{X}$, see [6, Lemma 1.1] or [4, Lemma 5.1.1]. Our first main result is, therefore an immediate consequence. By the equality of two norms we mean their pointwise equality.

Theorem 2.2. *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be rvs on $\mathbb{R}^d$, satisfying condition ($\mathcal{H}$), with dfs F and G. Then F = G if and only if $\|\cdot\|_F = \|\cdot\|_G$.*

In view of the above result we call every norm on $\mathbb{R}^{d+1}$ which has the representation (1) an *F-norm*.

Let us point out that Theorem 2.2 is still valid when $\boldsymbol{X}$ is not assumed to have nonzero components, but the mapping $\|\cdot\|_F$ is then actually only a *seminorm* on $\mathbb{R}^{d+1}$. Extending the definition of the max-CF of $\boldsymbol{X}$ by considering the mapping $\|\cdot\|_F$ thus generally leads to a seminorm rather than a norm. Observe though that unless $\boldsymbol{X}$ is the degenerate rv $\boldsymbol{0} \in \mathbb{R}^d$, the mapping $\|\cdot\|_F$ induces an F-norm on $\mathbb{R}^{d'+1}$, where d' is the number of nonzero components of $\boldsymbol{X}$. There is therefore no loss of generality in considering F-norms rather than F-seminorms, and we do so in the remainder of this paper.

An F-norm is usually conveniently calculated by using the following fundamental formula.

Lemma 2.3. *Let F be the df of a rv $\boldsymbol{X}$ satisfying condition ($\mathcal{H}$). Then, for any $\boldsymbol{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$, we have*

$$\|\boldsymbol{x}\|_F = |x_0| + \int_{|x_0|}^{\infty} [1 - F(t/|x_1|, \ldots, t/|x_d|)] \, dt$$

with the convention $1/0 = \infty$.

Proof. This is a straightforward consequence of the well-known formula

$$E(|Z|) = \int_{0}^{\infty} P(|Z| > t) \, dt$$

applied to the nonnegative rv $Z = \max(|x_0|, |x_1| X_1, \ldots, |x_d| X_d)$. $\qquad \square$

Example 2.1 (Degenerate F-norm)**.** The degenerate distribution concentrated at a d-dimensional vector $\boldsymbol{c} = (c_1, \ldots, c_d) > \boldsymbol{0}$ is characterized by the F-norm

$$\|\boldsymbol{x}\|_F = \max(|x_0|, c_1|x_1|, \ldots, c_d|x_d|).$$

In particular, the standard sup-norm $\|\boldsymbol{x}\|_\infty := \max_{0 \leq i \leq d} |x_i|$ on $\mathbb{R}^{d+1}$ is an F-norm which characterizes the constant rv $(1, \ldots, 1) \in \mathbb{R}^d$.

Example 2.2 (Bernoulli F-norm)**.** The Bernoulli distribution with parameter $p \in (0, 1)$ is characterized by the bivariate F-norm

$$\left\|(x_0, x_1)\right\|_F = (1 - p)|x_0| + p \max(|x_0|, |x_1|).$$

Example 2.3 (Uniform F-norm)**.** The uniform distribution on $(0, 1)$ is characterized by the bivariate F-norm

$$\left\|(x_0, x_1)\right\|_F = \begin{cases} |x_0|, & \text{if } |x_1| \leq |x_0|, \\ |x_0| + \displaystyle\int_{|x_0|}^{|x_1|} \left(1 - \frac{t}{|x_1|}\right) dt = \frac{x_0^2 + x_1^2}{2|x_1|}, & \text{if } |x_1| > |x_0|. \end{cases}$$

Example 2.4 (Exponential F-norm)**.** The exponential distribution with mean $1/\lambda$, $\lambda > 0$ is characterized by the bivariate F-norm

$$\left\|(x_0, x_1)\right\|_F = |x_0| + \int_{|x_0|}^{\infty} \exp\left(-\lambda \frac{t}{|x_1|}\right) dt = |x_0| + \frac{|x_1|}{\lambda} \exp\left(-\lambda \frac{|x_0|}{|x_1|}\right)$$

when $x_1 \neq 0$, and $|x_0|$ otherwise.

We now explore some simple properties of F-norms. Each F-norm induces, as a norm, a continuous function on $\mathbb{R}^{d+1}$. It takes the value 1 at $(1, 0, \ldots, 0)$. It also defines a radially symmetric function, *i.e.*

$$\forall \boldsymbol{x} \in \mathbb{R}^{d+1}, \ \|\boldsymbol{x}\|_F = \||\boldsymbol{x}|\|_F, \ \text{with} \ |\boldsymbol{x}| = (|x_0|, |x_1|, \ldots, |x_d|).$$

The norm $\|\cdot\|_F$ is, therefore, determined by its values on $[0, \infty)^{d+1}$. Additionally, any F-norm defines a monotone norm on $\mathbb{R}^{d+1}$ in the sense that

$$\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{y} \Rightarrow \|\boldsymbol{x}\|_F \leq \|\boldsymbol{y}\|_F.$$

These properties make it possible, in some cases, to show that certain norms are not F-norms:
- the norm $\|\cdot\| := 2 \|\cdot\|_\infty$ is not an F-norm because $\left\|(1, 0, \ldots, 0)\right\| = 2$,
- for any $\delta \in (0, 1)$, the matrix

$$M = \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}$$

is symmetric and positive definite, and therefore induces the norm

$$\left\|(x_0, x_1)\right\|_\delta := \left[(x_0, x_1) M (x_0, x_1)^\top\right]^{1/2} = \sqrt{x_1^2 - 2\delta x_1 x_2 + x_2^2}.$$

This norm is not radially symmetric, as

$$\left\|(1, -1)\right\|_\delta = \sqrt{2}\sqrt{1 + \delta} \neq \sqrt{2}\sqrt{1 - \delta} = \left\|(1, 1)\right\|_\delta.$$

It is actually not monotone either, since

$$(1, 0) \leq (1, \delta) \ \text{but} \ \left\|(1, 0)\right\|_\delta = 1 > \sqrt{1 - \delta^2} = \left\|(1, \delta)\right\|_\delta.$$

The norm $\|\cdot\|_\delta$ therefore cannot be an F-norm.

The proof of the following result, which provides bounds for a general F-norm, is elementary.

Proposition 2.4. *Let* $\boldsymbol{X} = (X_1, \ldots, X_d)$ *be a rv satisfying* $(\mathcal{H})$ *and* $\|\cdot\|_F$ *be the corresponding F-norm. For any* $\boldsymbol{x} \in \mathbb{R}^{d+1}$, *we have the bounds*

$$\max(|x_0|, |x_1| E(X_1), \ldots, |x_d| E(X_d)) \leq \|\boldsymbol{x}\|_F \leq |x_0| + \sum_{i=1}^{d} |x_i| E(X_i).$$

The upper bound is always strict if both x_0 and at least one of the x_i $(1 \leq i \leq d)$ are nonzero.

While the upper bound in Proposition 2.4 is not an F-norm, the weighted sup-norm in the lower bound is, see Example 2.1. In the case $E(X_1) = \cdots = E(X_d) = 1$, this is just the standard sup-norm on $\mathbb{R}^{d+1}$.

We close this section by providing results to identify those norms which are F-norms. Let us highlight first that for any norm $\|\cdot\|$ on $\mathbb{R}^{d+1}$ and any $\boldsymbol{x} \in \mathbb{R}^d$, the function $t \mapsto \|(t, \boldsymbol{x})\|$ is convex on $[0, \infty)$ (and right-continuous at 0), and therefore automatically absolutely continuous on this interval [see *e.g.* 12]. With this in mind, we have the following preliminary result.

Lemma 2.5. *A norm $\|\cdot\|$ on $\mathbb{R}^{d+1}$ is an F-norm if and only if the following two conditions hold:*
(i) it is radially symmetric,
(ii) there exists a rv $\boldsymbol{X} = (X_1, \ldots, X_d)$ which satisfies ($\mathcal{H}$) such that for any $x_1, \ldots, x_d > 0$, the Lebesgue derivative of $t \mapsto \|(t, 1/x_1, \ldots, 1/x_d)\|$ is equal to $P(X_1 \le tx_1, \ldots, X_d \le tx_d)$ almost everywhere, and

$$\left\| \left(0, \frac{1}{x_1}, \ldots, \frac{1}{x_d} \right) \right\| = E\left(\max\left(\frac{X_1}{x_1}, \ldots, \frac{X_d}{x_d} \right) \right).$$

In that case then $\|\cdot\| = \|\cdot\|_F$ with F being the df of $\boldsymbol{X}$.

Proof. That any F-norm satisfies (i) is obvious, while (ii) is a clear consequence of Lemma 2.3, reformulated as

$$\left\| (t, 1/x_1, \ldots, 1/x_d) \right\|_F = t + \int_t^\infty [1 - P(X_1 \le ux_1, \ldots, X_d \le ux_d)]\, du$$

when $\boldsymbol{X}$ has df F.

Conversely, let $\|\cdot\|$ satisfy (i) and (ii). Since $\|\cdot\|$ and $\|\cdot\|_F$ are continuous, as well as radially symmetric by (i), we only need to show that $\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_F$ for all $\boldsymbol{x} > \boldsymbol{0}$. Pick such an $\boldsymbol{x}$ and write it as $\boldsymbol{x} = (t, 1/x_1, \ldots, 1/x_d)$, for $t, x_1, \ldots, x_d > 0$. Write then, by absolute continuity,

$$\left\| (t, 1/x_1, \ldots, 1/x_d) \right\| - \left\| (0, 1/x_1, \ldots, 1/x_d) \right\|$$

$$= t - \int_0^t [1 - P(X_1 \le ux_1, \ldots, X_d \le ux_d)]\, du$$

$$= t + \int_t^\infty [1 - P(X_1 \le ux_1, \ldots, X_d \le ux_d)]\, du - E\left(\max\left(\frac{X_1}{x_1}, \ldots, \frac{X_d}{x_d} \right) \right).$$

Applying Lemma 2.3 and noting that by (ii),

$$\left\| \left(0, \frac{1}{x_1}, \ldots, \frac{1}{x_d} \right) \right\| = E\left(\max\left(\frac{X_1}{x_1}, \ldots, \frac{X_d}{x_d} \right) \right),$$

concludes the proof. □

Although the above result and in particular its part (ii) seem to be a tautology in view of the definition of an F-norm in (1), it turns out to be quite a powerful tool of its own as illustrated by Corollaries 2.6 and 2.7 below. We start by the following simple corollary in two dimensions.

Corollary 2.6. *A norm $\|\cdot\|$ on $\mathbb{R}^2$ is an F-norm if and only if the following two conditions hold:*
(i) it is radially symmetric,
(ii) the Lebesgue derivative of $t \mapsto \|(t, 1)\|$ is almost everywhere equal to a univariate df F on $[0, \infty)$ with a finite first moment equal to $\|(0, 1)\|$.

In that case then $\|\cdot\| = \|\cdot\|_F$.

Example 2.5 (On the L^1-norm)**.** The L^1-norm $\|(x_0, x_1)\| = |x_0| + |x_1|$ on $\mathbb{R}^2$ is not an F-norm. Indeed, we have

$$\frac{d}{dt}(\|(t, 1)\|) = 1, \ t > 0,$$

which does not define a df on $[0, \infty)$ having a (strictly) positive first moment.

Example 2.6 (On the L^p-norm). Each L^p-norm $\left\|(x_0, x_1)\right\|_p = (|x_0|^p + |x_1|^p)^{1/p}$ on $\mathbb{R}^2$, with $1 < p < \infty$, is an F-norm. Indeed, it is clearly radially symmetric and

$$\frac{d}{dt}\left(\left\|(t, 1)\right\|_p\right) = \frac{d}{dt}\left((t^p + 1)^{1/p}\right) = (1 + t^{-p})^{1/p-1}, \ t > 0,$$

which defines the df of a Burr type III distribution in the sense of [1, Table 2.1]. This distribution, for $p > 1$, has a finite first moment.

Even though providing a simple characterization of F-norms in arbitrary dimensions appears to be a difficult problem due to the high-level condition (ii) in Lemma 2.5, there is a simple inversion formula inspired by this result that makes it possible to go from an F-norm to its pertaining df. This is the focus of the following result, which can also be used to check that a norm is *not* an F-norm. Its proof is a straightforward consequence of Lemma 2.3 and right-continuity of the df F. The fact that the df F is determined by $\|\cdot\|_F$ is another obvious consequence.

Corollary 2.7. *Let $\|\cdot\|_F$ be an F-norm. Then, for any $x_1, \ldots, x_d > 0$, the right-derivative of the function $t \mapsto \left\|(t, 1/x_1, \ldots, 1/x_d)\right\|_F$ at $t = 1$ exists and is $F(x_1, \ldots, x_d)$.*

Example 2.7 (On the L^1-norm again). The L^1-norm

$$\left\|(x_0, x_1, \ldots, x_d)\right\| = \sum_{i=0}^{d} |x_i|$$

on $\mathbb{R}^{d+1}$ is not an F-norm. Indeed, we have, for any $x_1, \ldots, x_d > 0$,

$$\frac{d}{dt}(\left\|(t, 1/x_1, \ldots, 1/x_d)\right\|) = 1, \ t > 0,$$

which defines the df of the degenerate vector $(0, \ldots, 0)$. This distribution does not have strictly positive marginal moments and thus, by Corollary 2.7, $\|\cdot\|$ cannot be an F-norm.

3 Limiting behavior and estimation of F-norms

While the pointwise limit of a convergent sequence of D-norms is again a D-norm [see 4, Corollary 1.8.5], this is not true for F-norms: for instance, if (p_n) is a sequence of real numbers with $p_n > 1$ and $p_n \downarrow 1$, then $\|\cdot\|_{p_n} \to \|\cdot\|_1$, and $\|\cdot\|_{p_n}$ is for each n an F-norm, but the limit $\|\cdot\|_1$ is not.

However, if we ask that the limit is an F-norm, then we can relate the convergence of F-norms with convergence of distributions in the *Wasserstein metric*. Recall that the Wasserstein metric between two probability distributions P, Q on $\mathbb{R}^d$ with finite first moments in each component is

$$d_W(P, Q)$$
$$:= \ \inf\{E(\|\boldsymbol{X} - \boldsymbol{Y}\|_1): \ \boldsymbol{X} \text{ has distribution } P, \ \boldsymbol{Y} \text{ has distribution } Q\}.$$

Convergence of probability measures P_n to P on $\mathbb{R}^d$ with respect to the Wasserstein metric is equivalent to weak convergence together with convergence of the moments

$$\int_{\mathbb{R}^d} \|\boldsymbol{x}\|_1 \ P_n(d\boldsymbol{x}) \to \int_{\mathbb{R}^d} \|\boldsymbol{x}\|_1 \ P(d\boldsymbol{x});$$

see *e.g.* [16, Definition 6.8 and Theorem 6.9]. With this definition in mind, we can show the following result.

Theorem 3.1. *Pointwise convergence of a sequence of F-norms $\|\cdot\|_{F_n}$ to an F-norm $\|\cdot\|_F$ is equivalent to convergence of the sequence of distributions F_n to F in the Wasserstein metric.*

Proof. Pointwise convergence of $\|\cdot\|_{F_n}$ to $\|\cdot\|_F$ implies pointwise convergence of the sequence of max-CFs of F_n (as defined in (2)) to the max-CF of F, which entails the desired convergence in the Wasserstein metric by Theorem 2.1 in [6].

Conversely, if $F_n \to F$ in the Wasserstein metric, let $\boldsymbol{X}^{(n)}$ and $\boldsymbol{X}$ have dfs F_n and F. For any $\boldsymbol{x} = (x_0, x_1, \ldots, x_d) \geq \boldsymbol{0}$,

$$
\begin{aligned}
& \max(x_0, x_1 X_1^{(n)}, \ldots, x_d X_d^{(n)}) \\
= \ & \max(x_0, x_1[X_1 + (X_1^{(n)} - X_1)], \ldots, x_d[X_d + (X_d^{(n)} - X_d)]) \\
\leq \ & \max(x_0, x_1 X_1, \ldots, x_d X_d) + \max_{1 \leq i \leq d} x_i |X_i^{(n)} - X_i|.
\end{aligned}
$$

An analogue inequality holds if we switch $\boldsymbol{X}^{(n)}$ and $\boldsymbol{X}$. We can then integrate to find

$$
\left| \|\boldsymbol{x}\|_{F_n} - \|\boldsymbol{x}\|_F \right| \leq \|\boldsymbol{x}\|_\infty \, E\left(\left\| \boldsymbol{X}^{(n)} - \boldsymbol{X} \right\|_1 \right).
$$

Since $\boldsymbol{X}^{(n)}$ and $\boldsymbol{X}$ were arbitrary rvs having dfs F_n and F, this yields

$$
\left| \|\boldsymbol{x}\|_{F_n} - \|\boldsymbol{x}\|_F \right| \leq \|\boldsymbol{x}\|_\infty \, d_W(F_n, F) \to 0 \tag{3}
$$

which concludes the proof. □

The nice behavior of F-norms with respect to sequences of distributions naturally raises the question of what happens when F_n is chosen to be the empirical df based on iid copies $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)}$ of a rv $\boldsymbol{X}$ satisfying $(\mathcal{H})$, i.e.

$$
\widehat{F}_n(\boldsymbol{t}) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\boldsymbol{X}^{(i)} \leq \boldsymbol{t}\}}, \quad \boldsymbol{t} \in \mathbb{R}^d.
$$

The (random) F-norm generated by $\widehat{F}_n$ is nothing but

$$
\|\boldsymbol{x}\|_{\widehat{F}_n} = \frac{1}{n} \sum_{i=1}^n \max\left(|x_0|, |x_1| X_1^{(i)}, \ldots, |x_d| X_d^{(i)} \right).
$$

The law of large numbers then implies, for each $\boldsymbol{x} \in \mathbb{R}^{d+1}$, that a.s.

$$
\|\boldsymbol{x}\|_{\widehat{F}_n} \to E\left(\max\left(|x_0|, |x_1| X_1, \ldots, |x_d| X_d \right) \right) = \|\boldsymbol{x}\|_F \quad \text{as } n \to \infty.
$$

This convergence suggests that the estimation of an F-norm is completely straightforward; by contrast, estimating the related concept of a D-norm in the context of multivariate extreme value analysis requires quite sophisticated techniques.

We now provide further insight into the convergence of $\|\cdot\|_{\widehat{F}_n}$ to $\|\cdot\|_F$. Noting that for any $\boldsymbol{x}$ in a box $K = \prod_{i=0}^d [a_i, b_i] \subset [0, \infty)^{d+1}$ we have, by monotonicity of F-norms,

$$
\begin{aligned}
\|\boldsymbol{x}\|_{\widehat{F}_n} - \|\boldsymbol{x}\|_F \ & \leq \ \left(\|\boldsymbol{b}\|_{\widehat{F}_n} - \|\boldsymbol{b}\|_F \right) + \left(\|\boldsymbol{b}\|_F - \|\boldsymbol{a}\|_F \right) \\
\text{and} \ \ \|\boldsymbol{x}\|_F - \|\boldsymbol{x}\|_{\widehat{F}_n} \ & \leq \ \left(\|\boldsymbol{a}\|_F - \|\boldsymbol{a}\|_{\widehat{F}_n} \right) + \left(\|\boldsymbol{b}\|_F - \|\boldsymbol{a}\|_F \right),
\end{aligned}
$$

the following locally uniform refinement of the pointwise almost sure convergence of $\|\cdot\|_{\widehat{F}_n}$ to $\|\cdot\|_F$ is a direct consequence of the continuity of $\|\cdot\|_F$.

Theorem 3.2. *Let $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)}$ be iid copies of a rv $\boldsymbol{X}$ satisfying $(\mathcal{H})$, with df F. Let $\|\cdot\|_{\widehat{F}_n}$ be the random F-norm generated by the empirical df $\widehat{F}_n$ of this sample. We then have, for any $\boldsymbol{x}_0 \geq \boldsymbol{0}$,*

$$
\sup_{\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{x}_0} \left| \|\boldsymbol{x}\|_{\widehat{F}_n} - \|\boldsymbol{x}\|_F \right| \to 0 \ \text{ a.s.}
$$

To analyse the rate of (uniform) convergence of $\|\cdot\|_{\widehat{F}_n}$ to $\|\cdot\|_F$, we define the *empirical F-norm process*

$$S_n = (S_n(\boldsymbol{x}))_{\boldsymbol{x} \geq \boldsymbol{0}} := \sqrt{n} \left(\|\boldsymbol{x}\|_{\widehat{F}_n} - \|\boldsymbol{x}\|_F \right)_{\boldsymbol{x} \geq \boldsymbol{0}}$$

on $[0, \infty)^{d+1}$. This stochastic process has continuous sample paths and satisfies $S_n(\boldsymbol{0}) = 0$. Suppose then that $E(X_i^2) < \infty$ for any $i \in \{1, \dots, d\}$. Based on the standard central limit theorem, which gives the pointwise asymptotic normality of S_n, we may ask the question of the limiting behavior of the process S_n. For ease of exposition, we state a result in the case $d = 1$.

Theorem 3.3. *Let $X^{(1)}, \dots, X^{(n)}$ be iid copies of a univariate rv X with df F. Assume that X is nonnegative, with nonzero expectation and finite variance. Let $\|\cdot\|_{\widehat{F}_n}$ be the random F-norm generated by the empirical df $\widehat{F}_n$ of this sample. For any $x_0, y_0 > 0$, we have*

$$S_n(x, y) := \sqrt{n} \left(\|(x, y)\|_{\widehat{F}_n} - \|(x, y)\|_F \right) \to S(x, y)$$

weakly in the space $C([0, x_0] \times [0, y_0])$ of continuous functions over $[0, x_0] \times [0, y_0]$, where the limiting process S, which should be read as 0 when $y = 0$, is a bivariate Gaussian process with covariance structure

$$\text{Cov}(S(x_1, y_1), S(x_2, y_2))$$
$$= x_1 x_2 \iint\limits_{[1, \infty)^2} \left[F\left(\min\left\{ \frac{x_1}{y_1} u, \frac{x_2}{y_2} v \right\} \right) - F\left(\frac{x_1}{y_1} u \right) F\left(\frac{x_2}{y_2} v \right) \right] du \, dv.$$

Under the further assumption that $\int_0^\infty \sqrt{F(u)[1 - F(u)]} \, du < \infty$ [which is equivalent to $E(X^2) < \infty$ when F is regularly varying at infinity, according to *e.g.* 13, p.276] we have the representation

$$S(x, y) \overset{d}{=} y \int\limits_{x/y}^{\infty} W \circ F(u) \, du$$

as processes in $C([0, x_0] \times [0, y_0])$, where W is a Brownian bridge on $[0, 1]$. Indeed, since for any $t \in [0, 1]$ the rv $W(t)$ is Gaussian centered with variance $t(1 - t)$, we have $E|W(t)| = \sqrt{t(1 - t)}\sqrt{2/\pi}$, and thus

$$E \left| \int\limits_{x/y}^{\infty} W \circ F(u) \, du \right| \quad \leq \quad \int\limits_0^{\infty} E|W \circ F(u)| \, du$$

$$= \quad \sqrt{\frac{2}{\pi}} \int\limits_0^{\infty} \sqrt{F(u)[1 - F(u)]} \, du < \infty$$

so that $y \int_{x/y}^{\infty} W \circ F(u) \, du$ is well-defined and finite with probability 1. It is then straightforward to show, using the covariance properties of W, that the covariance structure of this Gaussian process coincides with that of S.

Proof. The random functions S_n and S are elements of the functional space $C([0, x_0] \times [0, y_0])$. By Theorem 7.5 in [2], it suffices to show the convergence of finite-dimensional margins of S_n to those of S along with tightness of (S_n), in the sense of tightness of its sequence of distributions.

We start by convergence of finite-dimensional margins. The multivariate central limit theorem implies, for nonnegative pairs $(x_1, y_1), \dots, (x_k, y_k)$, that the rv $(S_n(x_1, y_1), \dots, S_n(x_k, y_k))$ converges weakly to a centered Gaussian distribution. By Hoeffding's identity [see 4, Lemma 2.5.2], the limiting covariance matrix is described by

$$\text{Cov}(\max(x_i, y_i X), \max(x_j, y_j X))$$
$$= \iint\limits_{\mathbb{R}^2} [P(\max(x_i, y_i X) \leq x, \max(x_j, y_j X) \leq y)$$

$$- P(\max(x_i, y_i X) \le x) P(\max(x_j, y_j X) \le y)]\, dx\, dy$$

$$= \int_{x_i}^{\infty} \int_{x_j}^{\infty} [P(y_i X \le x, y_j X \le y) - P(y_i X \le x) P(y_j X \le y)]\, dx\, dy.$$

This is clearly equal to 0 when either y_i or y_j is 0, and otherwise, using the change of variables $x = x_i u$, $y = x_j v$, we find

$$\mathrm{Cov}(\max(x_i, y_i X), \max(x_j, y_j X))$$

$$= x_i x_j \iint_{[1,\infty)^2} \left[F\left(\min\left\{ \frac{x_i}{y_i} u, \frac{x_j}{y_j} v \right\} \right) - F\left(\frac{x_i}{y_i} u \right) F\left(\frac{x_j}{y_j} v \right) \right]\, du\, dv$$

which is exactly the covariance structure of the Gaussian process S.

We now show tightness, that is, for any $\varepsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P\left(\sup_{\substack{(x_1, y_1),(x_2,y_2) \in [0,x_0] \times [0,y_0] \\ \max(|x_1 - x_2|, |y_1 - y_2|) \le \delta}} |S_n(x_1, y_1) - S_n(x_2, y_2)| > \varepsilon \right) = 0,$$

or, in other words, that S_n is stochastically equicontinuous on $[0, x_0] \times [0, y_0]$. The key to the proof is three-fold. Firstly, we apply Theorem 1 p.93 of [15] to construct, on a common probability space, a triangular array $(U^{(n,1)}, \dots, U^{(n,n)})_{n \ge 1}$ of rowwise independent, standard uniform rvs, and a Brownian bridge $\widetilde{W}$ such that

$$\sup_{0 \le t \le 1} |\mathbb{W}_n(t) - \widetilde{W}(t)| \to 0 \text{ a.s. with } \mathbb{W}_n(t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\mathbb{1}_{\{U^{(n,i)} \le t\}} - t \right].$$

Secondly, if we denote by q the quantile function of X (i.e. the left-continuous inverse of F) and by $\widetilde{X}^{(n,i)} := q(U^{(n,i)})$, we have, for any $n \ge 1$,

$$S_n(x, y) \stackrel{d}{=} \widetilde{S}_n(x, y) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\max\left(x, y \widetilde{X}^{(n,i)} \right) - E(\max(x, yX)) \right],$$

as processes in $C([0, x_0] \times [0, y_0])$. We may and will therefore prove our result using $\widetilde{S}_n$ rather than S_n. Thirdly and finally, if $(x, y) \in [0, x_0] \times [0, y_0]$, we have

$$\min\left(x, y \widetilde{X}^{(n,i)} \right) = y \int_0^{x/y} \left[1 - \mathbb{1}_{\{\widetilde{X}^{(n,i)} \le u\}} \right]\, du.$$

Since $\widetilde{X}^{(n,i)} \le u \Leftrightarrow U^{(n,i)} \le F(u)$, this yields

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\min\left(x, y \widetilde{X}^{(n,i)} \right) - E(\min(x, yX)) \right] = -y \int_0^{x/y} \mathbb{W}_n \circ F(u)\, du.$$

Using the identity $\max(a, b) = a + b - \min(a, b)$, valid for any $a, b \ge 0$, it follows that:

$$\widetilde{S}_n(x_1, y_1) - \widetilde{S}_n(x_2, y_2) = (y_1 - y_2) \times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\widetilde{X}^{(n,i)} - E(X)]$$

$$+ T_n(x_1, y_1) - T_n(x_2, y_2)$$

$$\text{with } T_n(x, y) := y \int_0^{x/y} \mathbb{W}_n \circ F(u)\, du.$$

The first term on the right-hand side above is stochastically equicontinuous, because the random term is a $O_P(1)$ (by the Chebyshev inequality). We conclude the proof by focusing on $T_n(x, y)$, and for this we first remark that

$$\sup_{\substack{0 \le x \le x_0 \\ 0 \le y \le y_0}} \left| y \int_0^{x/y} \left[\mathbb{W}_n \circ F(u) - \widetilde{W} \circ F(u) \right] du \right| \le x_0 \sup_{0 \le t \le 1} |\mathbb{W}_n(t) - \widetilde{W}(t)| \to 0$$

almost surely. A consequence of this convergence is that, to show the stochastic equicontinuity of T_n, it is enough to prove that the random function T defined by

$$T(x, y) := \begin{cases} y \int_0^{x/y} \widetilde{W} \circ F(u) \, du & \text{if } y > 0, \\ 0 & \text{if } y = 0, \end{cases}$$

satisfies

$$\lim_{\delta \to 0} P \left(\sup_{\substack{(x_1, y_1), (x_2, y_2) \in [0, x_0] \times [0, y_0] \\ \max(|x_1 - x_2|, |y_1 - y_2|) \le \delta}} |T(x_1, y_1) - T(x_2, y_2)| > \varepsilon \right) = 0.$$

Recall that $\widetilde{W}$ has almost surely continuous sample paths on $[0, 1]$, and thus T is almost surely continuous on $[0, x_0] \times (0, y_0]$. Because, for any $y > 0$,

$$T(x, y) = \int_0^x \widetilde{W} \circ F(u/y) \, du$$

and $T(x, 0) = 0$, it follows by the dominated convergence theorem that almost sure continuity of T also holds on the *compact* set $[0, x_0] \times [0, y_0]$. Then T must also be almost surely uniformly continuous on this set, and therefore

$$\lim_{\delta \to 0} \sup_{\substack{(x_1, y_1), (x_2, y_2) \in [0, x_0] \times [0, y_0] \\ \max(|x_1 - x_2|, |y_1 - y_2|) \le \delta}} |T(x_1, y_1) - T(x_2, y_2)| = 0 \text{ a.s.}$$

This completes the proof.   □

In the case $d > 1$, and under regularity conditions [*e.g.* those of 10], a similar proof using a special construction of the multivariate empirical process can be written to show an analogue of Theorem 3.3, which gives the convergence of the process S_n, in a space of continuous functions over compact subsets of $[0, \infty)^{d+1}$, to a $(d + 1)$–dimensional Gaussian process S with covariance structure

$$\text{Cov}(S(x_1, \boldsymbol{x}_1), S(x_2, \boldsymbol{x}_2))$$
$$= x_1 x_2 \iint_{[1, \infty)^{2d}} \left[F \left(\min \left\{ \frac{x_1}{\boldsymbol{x}_1} \boldsymbol{u}, \frac{x_2}{\boldsymbol{x}_2} \boldsymbol{v} \right\} \right) - F \left(\frac{x_1}{\boldsymbol{x}_1} \boldsymbol{u} \right) F \left(\frac{x_2}{\boldsymbol{x}_2} \boldsymbol{v} \right) \right] d\boldsymbol{u} \, d\boldsymbol{v}.$$

Our objective is now to dwell upon the nice sequential behavior of F-norms and show an example of how this could be used to prove powerful theorems on the convergence of certain sequences of rvs. To this end we first need to understand better how to manipulate F-norms, which leads us to exploring their algebraic properties.

4 Algebra of the set of F-norms

One can multiply F-norms $\|\cdot\|_F$ and $\|\cdot\|_G$ by constructing the F-norm generated by the componentwise product of any pair of *independent* rvs having dfs F and G; independence is used to ensure that the distribution of this

componentwise product is well-defined, and thus so is the product F-norm. We denote this operation by $\|\cdot\|_F \star \|\cdot\|_G$. It coincides with taking products of D-norms if $\|\cdot\|_F$ and $\|\cdot\|_G$ have components with expectation 1, see [4, Section 1.9]. In general, the product F-norm can be calculated using the following Tonelli formula.

Proposition 4.1. *Let F and G be the dfs of two rvs satisfying condition $(\mathcal{H})$. Then, for any $\boldsymbol{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$,*

$$
\begin{aligned}
(\|\cdot\|_F \star \|\cdot\|_G)(\boldsymbol{x}) &= \int_{[0,\infty)^d} \big\|(x_0, x_1 t_1, \ldots, x_d t_d)\big\|_F \, dG(t_1, \ldots, t_d) \\
&= \int_{[0,\infty)^d} \big\|(x_0, x_1 t_1, \ldots, x_d t_d)\big\|_G \, dF(t_1, \ldots, t_d).
\end{aligned}
$$

Proof. Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ be independent and have dfs F and G. We have

$$
(\|\cdot\|_F \star \|\cdot\|_G)(\boldsymbol{x}) = E\left(\max\left(|x_0|, |x_1| X_1 Y_1, \ldots, |x_d| X_d Y_d\right)\right).
$$

By nonnegativity of $\max(|x_0|, |x_1| X_1 Y_1, \ldots, |x_d| X_d Y_d)$ and independence of $\boldsymbol{X}$ and $\boldsymbol{Y}$, we find, using the Tonelli theorem, that

$$
\begin{aligned}
&(\|\cdot\|_F \star \|\cdot\|_G)(\boldsymbol{x}) \\
&= \int_{[0,\infty)^d} E\left(\max\left(|x_0|, |x_1| t_1 X_1, \ldots, |x_d| t_d X_d\right)\right) dG(t_1, \ldots, t_d)
\end{aligned}
$$

which is exactly the first formula. The second expression follows by swapping integration with respect to dG for integration with respect to dF. □

Example 4.1 (Product of uniform F-norms). Following Example 2.3, the product of the standard uniform F-norm by itself has the expression

$$
\begin{aligned}
(\|\cdot\|_F \star \|\cdot\|_F)(x_0, x_1) &= \int_0^1 \left(|x_0| \mathbb{1}_{\{t|x_1| \le |x_0|\}} + \frac{x_0^2 + t^2 x_1^2}{2t|x_1|} \mathbb{1}_{\{t|x_1| > |x_0|\}}\right) dt \\
&= \begin{cases} |x_0| & \text{if } |x_1| \le |x_0|, \\ \dfrac{5}{4}|x_1| + \dfrac{x_0^2}{2|x_1|}\left[\log\left(\dfrac{|x_1|}{|x_0|}\right) - \dfrac{1}{2}\right] & \text{if } |x_1| > |x_0|. \end{cases}
\end{aligned}
$$

Let us now explore in more detail the structure of the set of F-norms equipped with its multiplication. It is clear that the sup-norm $\|\cdot\|_\infty$ on $\mathbb{R}^{d+1}$, with generator $(1, \ldots, 1) \in \mathbb{R}^d$, is an identity element for this operation. It is also straightforward to see that it is the unique such element: if $\|\cdot\|_F$ is an identity element for $\star$ then $\|\cdot\|_F = \|\cdot\|_F \star \|\cdot\|_\infty = \|\cdot\|_\infty$. We summarize this short discussion by the following result.

Proposition 4.2. *The set of F-norms is a commutative monoid for the F-norm multiplication $\star$, with identity element $\|\cdot\|_\infty$. The only invertible elements are the F-norms generated by nonrandom vectors.*

The only point we need to show in Proposition 4.2 is the assertion about invertible elements. The key is to note the following lemmas.

Lemma 4.3. *Let Z be a real-valued rv such that $|E(e^{itZ})| = 1$ for any $t \in \mathbb{R}$. Then Z is almost surely constant.*

Proof of Lemma 4.3. We use the Cauchy-Schwarz inequality for the inner product $(X, Y) \mapsto E(X\overline{Y})$ on the space of complex-valued square-integrable rvs, to obtain:

$$
\forall t \in \mathbb{R}, \ |E(e^{itZ})|^2 = |E(e^{itZ} \cdot 1)|^2 \le 1.
$$

By assumption, we actually have equality here. This means that for any t, the rvs e^{itZ} and 1 are almost surely proportional, *i.e.* $e^{itZ} = \lambda(t)$, with $\lambda(t) \in \mathbb{C}$. Define now the event $E_t := \{e^{itZ} = \lambda(t)\}$, and let (t_n), (t'_n) be two sequences converging to 0. Define $E = (\bigcap_n E_{t_n}) \cap (\bigcap_n E_{t'_n})$. Then $P(E) = 1$ and on E,

$$\frac{\lambda(t_n) - \lambda(0)}{t_n} = \frac{e^{it_n Z} - 1}{t_n} \to iZ \text{ and } \frac{\lambda(t'_n) - \lambda(0)}{t'_n} = \frac{e^{it'_n Z} - 1}{t'_n} \to iZ.$$

It follows that the limit

$$\lim_{n \to \infty} \frac{\lambda(t_n) - \lambda(0)}{t_n}$$

exists and does not depend on the choice of $t_n \to 0$: the function λ is differentiable at 0. Conclude, by using (t_n) again, that on the event $(\bigcap_n E_{t_n})$, $\lambda'(0) = iZ$ and thus Z is almost surely the constant $-i\lambda'(0)$. □

Lemma 4.4. *Let X and Y be two independent nonnegative rvs such that $XY = 1$ almost surely. Then X and Y are almost surely positive constants.*

Proof of Lemma 4.4. Necessarily $P(X = 0) = P(Y = 0) = 0$. Then by assumption $\log X + \log Y$ is a.s. zero. Denote by $\varphi_X(t) := E(e^{it \log X})$ and $\varphi_Y(t) := E(e^{it \log Y})$ the characteristic functions of $\log X$ and $\log Y$. This entails $\varphi_X(t) \cdot \varphi_Y(t) = 1$ for any $t \in \mathbb{R}$, by independence. Since any characteristic function has a modulus at most 1, we find $|\varphi_X(t)| = |\varphi_Y(t)| = 1$. Conclude by applying Lemma 4.3. □

Proof of Proposition 4.2. Let $\|\cdot\|_F$ and $\|\cdot\|_G$ satisfy $\|\cdot\|_F \star \|\cdot\|_G = \|\cdot\|_\infty$. Equivalently, there are independent rvs $(X_1, \ldots, X_d)$ with df F and $(Y_1, \ldots, Y_d)$ with df G such that for any $\boldsymbol{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$,

$$E(\max(|x_0|, |x_1| X_1 Y_1, \ldots, |x_d| X_d Y_d)) = E(\max(|x_0|, |x_1| \cdot 1, \ldots, |x_d| \cdot 1)).$$

By Theorem 2.2, we find that each $X_i Y_i$ is a.s. constant equal to 1. Conclude by applying Lemma 4.4. □

The same kind of argument can be used to identify the set of idempotent elements for the multiplication of F-norms.

Proposition 4.5. *The only idempotent element for multiplication of F-norms is the sup-norm $\|\cdot\|_\infty$.*

The proof is again based on an auxiliary result for real-valued rvs.

Lemma 4.6. *Let X and Y be two independent nonnegative rvs having the same distribution and satisfying $XY \stackrel{d}{=} X$. Then $X = Y = 1$ almost surely.*

Proof of Lemma 4.6. The assumption is $P(XY \le t) = P(X \le t)$ for any t. Note that

$$
\begin{aligned}
P(X = 0) = P(XY = 0) &= P(X = 0) + P(Y = 0) - P(X = Y = 0) \\
&= P(X = 0)[2 - P(X = 0)]
\end{aligned}
$$

so that $P(X = 0) \in \{0, 1\}$, and necessarily $P(X = 0) = 0$ since $E(X) > 0$. Then by assumption $\log X + \log Y \stackrel{d}{=} \log X$. If $\varphi(t) := E(e^{it \log X})$ denotes the characteristic function of $\log X$, this entails $[\varphi(t)]^2 = \varphi(t)$ for any $t \in \mathbb{R}$, by independence. Thus, for any $t \in \mathbb{R}$, $\varphi(t) \in \{0, 1\}$. Noting that $\varphi(0) = 1$ and φ is continuous entails that necessarily $\varphi \equiv 1$, since $\varphi(\mathbb{R})$ must be a path-connected subset of $\{0, 1\}$. As a consequence, $\log X = 0$ almost surely, completing the proof. □

Proof of Proposition 4.5. Let $\|\cdot\|_F$ be an idempotent element for the multiplication of F-norms, with generator $(X_1, \ldots, X_d)$. Let $(Y_1, \ldots, Y_d)$ be an independent copy of this rv. Since $\|\cdot\|_F$ is idempotent, we have, for any $\boldsymbol{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$,

$$E(\max(|x_0|, |x_1| X_1 Y_1, \ldots, |x_d| X_d Y_d)) = E(\max(|x_0|, |x_1| X_1, \ldots, |x_d| X_d)).$$

By Theorem 2.2, we find that $X_i Y_i \stackrel{d}{=} X_i$, for each $i \in \{1, \ldots, d\}$. Conclude by applying Lemma 4.6. □

5 *F*-norms of general random vectors

The concept of *F*-norms focuses on the distribution of an arbitrary multivariate rv with nonnegative and integrable components. Our purpose here is to show how we can also define, in a sensible way, a concept of *F*-norms for a rv whose components can attain negative values, under an integrability condition.

Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be an arbitrary rv satisfying $E(\exp(X_i)) < \infty$, $1 \le i \le d$. Then $\boldsymbol{Y} := \exp(\boldsymbol{X}) = (\exp(X_1), \ldots, \exp(X_d))$ generates an *F*-norm $\|\cdot\|_F^{\exp}$. As the function $x \mapsto \exp(x)$ is a bijection from the real line onto the interval $(0, \infty)$, the distribution of $\boldsymbol{X}$ is characterized by the *F*-norm $\|\cdot\|_F^{\exp}$, which we call a *log F-norm*.

Example 5.1 (Normal distribution). Put $Z := \exp(X - \sigma^2/2)$, where X follows the univariate normal distribution $N(0, \sigma^2)$. The rv Z is log-normal distributed with $E(Z) = 1$. The log *F*-norm of $X - \sigma^2/2$ is then just a *D*-norm and equals, for $x, y > 0$,

$$\|(x, y)\|_F^{\exp} = E\left(\max(x, yZ)\right) = x\Phi\left(\frac{\sigma}{2} + \frac{\log(x/y)}{\sigma}\right) + y\Phi\left(\frac{\sigma}{2} + \frac{\log(y/x)}{\sigma}\right),$$

which is the so-called *Hüsler-Reiss D*-norm with parameter $\sigma^2 > 0$ [see 4]; by Φ we denote the df of the standard normal distribution on $\mathbb{R}$. As a consequence, the normal distribution $N(-\sigma^2/2, \sigma^2)$ of $X - \sigma^2/2$ is characterized by the preceding norm $\|\cdot\|_F^{\exp}$.

More generally, the log *F*-norm of the normal distribution $N(\mu, \sigma^2)$ with arbitrary $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ is, for $x, y > 0$,

$$\|(x, y)\|_F^{\exp} = E\left(\max\left(x, y\exp\left(\mu + \frac{\sigma^2}{2}\right)Z\right)\right)$$
$$= x\Phi\left(\frac{\log(x/y) - \mu}{\sigma}\right) + y\exp\left(\mu + \frac{\sigma^2}{2}\right)\Phi\left(\sigma + \frac{\log(y/x) + \mu}{\sigma}\right).$$

By Corollary 2.6, we should find back the log-normal df from this *F*-norm by differentiating $\|(t, 1)\|_F^{\exp}$ on $(0, \infty)$. Clearly

$$\frac{d}{dt}(\|(t, 1)\|_F^{\exp}) = \Phi\left(\frac{\log(t) - \mu}{\sigma}\right) + \frac{1}{\sigma}\Phi'\left(\frac{\log(t) - \mu}{\sigma}\right)$$
$$- \frac{1}{t\sigma}\exp\left(\mu + \frac{\sigma^2}{2}\right)\Phi'\left(\sigma - \frac{\log(t) - \mu}{\sigma}\right).$$

Note also that

$$\Phi'\left(\sigma - \frac{\log(t) - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left[\sigma - \frac{\log(t) - \mu}{\sigma}\right]^2\right)$$
$$= t\exp\left(-\mu - \frac{\sigma^2}{2}\right) \times \Phi'\left(\frac{\log(t) - \mu}{\sigma}\right)$$

to find, as expected:

$$\frac{d}{dt}(\|(t, 1)\|_F^{\exp}) = \Phi\left(\frac{\log(t) - \mu}{\sigma}\right).$$

Combining the discussion we have developed in the previous example with Theorem 3.1 leads, without any further calculation, to the following immediate result. This serves as a further example of how the asymptotic results in Section 3 may be used to establish asymptotic theory.

Corollary 5.1. *Let (μ_n), (σ_n) be real-valued sequences such that $\mu_n \to \mu$ and $\sigma_n \to \sigma > 0$. Then:*
- *The sequence of log-normal distributions with parameters μ_n and σ_n^2 converges to the log-normal distribution with parameters μ and σ^2 in the Wasserstein metric.*

- *The sequence G_n of normal distributions with parameters μ_n and σ_n^2 converges in distribution to the normal distribution G with parameters μ and σ^2, and the moments of G_n converge to those of G.*

More generally, if $\boldsymbol{X}$ follows a multivariate normal distribution $N(\boldsymbol{\mu}, \Sigma)$ with mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and covariance matrix $\Sigma = (\sigma_{ij})_{1 \le i, j \le d}$, then each component $Y_i = \exp(X_i)$ is log-normal distributed with mean $E(Y_i) = \exp(\mu_i + \sigma_{ii}/2)$. In analogy to the D-norm generated by the normalized rv $\boldsymbol{Z} = \boldsymbol{Y}/E(\boldsymbol{Y})$ and called a *Hüsler-Reiss D-norm* [see 4], we call the F-norm corresponding to $\boldsymbol{Y}$ a *Hüsler-Reiss F-norm*. It characterizes the normal distribution $N(\boldsymbol{\mu}, \Sigma)$.

The concept of log F-norms for rvs with an arbitrary sign is not adapted solely to Gaussian distributions, as we show in the following examples.

Example 5.2 (Gumbel distribution). Let X have the standard negative Gumbel distribution, *i.e.*

$$P(X \le t) = 1 - e^{-e^t}, \ t \in \mathbb{R}.$$

Then $\exp(X)$ has a unit exponential distribution, and therefore the log F-norm characterizing the standard negative Gumbel distribution is

$$\left\| (x_0, x_1) \right\|_F^{\exp} = |x_0| + |x_1| \exp\left(-\frac{|x_0|}{|x_1|} \right)$$

when $x_1 \ne 0$, and $|x_0|$ otherwise (see Example 2.4).

Example 5.3 (On the central limit theorem). Let $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \dots$ be iid copies of a centered rv $\boldsymbol{X} = (X_1, \dots, X_d)$ having covariance matrix Σ, and a finite moment generating function in a neighborhood of the origin, *i.e.* there exists $\varepsilon > 0$ with $\varphi_j(t) := E(\exp(tX_j)) < \infty$ for any $|t| < \varepsilon$ and $1 \le j \le d$. The multivariate central limit theorem and continuous mapping theorem imply

$$\exp\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{X}^{(i)} \right) \xrightarrow{d} \exp(\boldsymbol{\xi}), \tag{4}$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_d)$ follows a multivariate normal distribution with mean vector zero and covariance matrix Σ. Besides, we have

$$E\left[\exp\left(\frac{2}{\sqrt{n}} \sum_{i=1}^n X_j^{(i)} \right) \right] = E\left[\prod_{i=1}^n \exp\left(\frac{2}{\sqrt{n}} X_j^{(i)} \right) \right] = \left[\varphi_j(2/\sqrt{n}) \right]^n.$$

Since X_j is centered with variance Σ_{jj}, we have by a Taylor expansion

$$E\left[\exp\left(\frac{2}{\sqrt{n}} \sum_{i=1}^n X_j^{(i)} \right) \right] = \left[1 + \frac{2}{n} \Sigma_{jj} + o\left(\frac{1}{n} \right) \right]^n \to e^{2\Sigma_{jj}}.$$

It follows that the sequence

$$\exp\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_j^{(i)} \right), \ n \ge 1,$$

has a bounded second moment and thus is *uniformly integrable* [see *e.g.* 2] for each $j = 1, \dots, d$. This entails convergence of the sequence of its first moments and, combined with (4) and Theorem 3.1, pointwise convergence of the generated log F-norms, *i.e.*

$$E\left(\max\left(x_0, x_1 \exp\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_1^{(i)} \right), \dots, x_d \exp\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_d^{(i)} \right) \right) \right)$$
$$\to E\left(\max\left(x_0, x_1 \exp(\xi_1), \dots, x_d \exp(\xi_d) \right) \right) \ \text{as} \ n \to \infty$$

for each $x_0, x_1, \dots, x_d \ge 0$. We thus have a convergence of F-norms akin to the central limit theorem.

We could, of course, have used in place of the exponential function any one-to-one increasing transformation from $\mathbb{R}$ to $(0, \infty)$ in order to define an F-norm for general rvs. The exponential function, however, interacts well with our notion of product of F-norms, in the sense that

$$\|\cdot\|_{\boldsymbol{X}}^{\exp} \star \|\cdot\|_{\boldsymbol{Y}}^{\exp} = \|\cdot\|_{\boldsymbol{X}+\boldsymbol{Y}}^{\exp}$$

if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent: a product of two log F-norms is the log F-norm corresponding to the convolution of their individual distributions.

6 Geometry of F-norms

The original motivation for constructing F-norms was to combine the distributional properties of a max-CF with the structure of a D-norm into a single mathematical object. We have so far concentrated on the information that F-norms bring about multivariate distributions. We use here the geometry of the F-norms to find yet other different objects who summarize a multivariate distribution.

Since any norm $\|\cdot\|$ on $\mathbb{R}^{d+1}$ is homogeneous, an immediate consequence is that each F-norm $\|\cdot\|_F$ is uniquely determined by its values on the unit sphere for $\|\cdot\|$, namely $S_{\|\cdot\|} := \left\{ \boldsymbol{u} \in \mathbb{R}^{d+1} : \|\boldsymbol{u}\| = 1 \right\}$: to put it differently, we have for $\boldsymbol{x} \in \mathbb{R}^{d+1}$, $\boldsymbol{x} \neq \boldsymbol{0}$,

$$\|\boldsymbol{x}\|_F = \|\boldsymbol{x}\| \left\| \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \right\|_F , \tag{5}$$

with $\boldsymbol{x} / \|\boldsymbol{x}\| \in S_{\|\cdot\|}$. By choosing $\|\cdot\| = \|\cdot\|_1$ and using the radial symmetry of the L^1-norm and of F-norms, we find that we need only consider the values of $\|\cdot\|_F$ on the part of the sphere $S_{\|\cdot\|}$ contained in $[0, \infty)^{d+1}$. In other words, each df F of a rv $\boldsymbol{X}$ satisfying $(\mathcal{H})$ is characterized by the function

$$A(\boldsymbol{t}) := \left\| \left(1 - \sum_{i=1}^{d} t_i, t_1, \ldots, t_d \right) \right\|_F , \quad \boldsymbol{t} = (t_1, \ldots, t_d),$$

defined on $\Delta_1 := \left\{ \boldsymbol{t} \in [0, 1]^d : \sum_{i=1}^{d} t_i \leq 1 \right\}$. This construction is similar to that of the *Pickands dependence function* in multivariate extreme value theory [see *e.g.* 8], and we therefore call the function A the Pickands dependence function of the F-norm $\|\cdot\|_F$. Let us briefly mention here that, based on a sample of copies of $\boldsymbol{X}$, we can estimate this Pickands dependence function by an empirical version, just as we did in Section 3 for the full F-norm: let $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)}$ be iid copies of a rv $\boldsymbol{X}$ satisfying $(\mathcal{H})$. Put, for $\boldsymbol{t} \in \Delta_1$, with $t_0 := 1 - \sum_{i=1}^{d} t_i$,

$$\widehat{A}_n(\boldsymbol{t}) := \frac{1}{n} \sum_{i=1}^{n} \max \left(t_0, t_1 X_1^{(i)}, \ldots, t_d X_d^{(i)} \right) ,$$

which is that (random) Pickands dependence function which characterizes the empirical df $\widehat{F}_n$. The asymptotic properties of $\widehat{A}_n$ follow directly from our asymptotic results in Section 3: since Δ_1 is compact, we get, by Theorem 3.2,

$$\sup_{\boldsymbol{t} \in \Delta_1} \left| \widehat{A}_n(\boldsymbol{t}) - A(\boldsymbol{t}) \right| \to 0 \text{ a.s.,}$$

and, by the multivariate extension of Theorem 3.3 mentioned at the end of Section 3, we have

$$\sqrt{n} \left(\widehat{A}_n(\boldsymbol{t}) - A(\boldsymbol{t}) \right) \to \mathcal{S}(\boldsymbol{t})$$

weakly in the space of the continuous functions on the unit simplex in $\mathbb{R}^{d+1}$, where $\mathcal{S}$ is a Gaussian process.

We now explore how, instead of characterizing an F-norm by a function such as its Pickands dependence function, we can identify it by a compact set which summarizes the geometry of an F-norm. Recall that an F-norm is characterized by its values on any sphere $S_{\|\cdot\|}$, where $\|\cdot\|$ is an arbitrary norm on $\mathbb{R}^{d+1}$. By choosing $\|\cdot\| = \|\cdot\|_F$ and using the radial symmetry of any F-norm, we obtain the following corollary.

Corollary 6.1. *Each F-norm $\|\cdot\|_F$ on $\mathbb{R}^{d+1}$ is characterized by the part of its unit sphere contained in the positive orthant of $\mathbb{R}^{d+1}$, that is:*

$$S^+_{\|\cdot\|_F} := S_{\|\cdot\|_F} \cap [0, \infty)^{d+1} = \left\{ \boldsymbol{x} \ge \boldsymbol{0} \in \mathbb{R}^{d+1} : \|\boldsymbol{x}\|_F = 1 \right\}.$$

This corollary provides a compact set characterizing any multivariate distribution with nonnegative, nonzero and integrable components. For such distributions, it is therefore an alternative to the lift zonoid studied by [9] and [11]. The next two examples show how this set can be computed in practice.

Example 6.1 (Unit sphere for the uniform F-norm). Let F be the uniform distribution on $(0, 1)$. We know from Example 2.3 that this distribution is characterized by the bivariate F-norm given by

$$\forall x_0, x_1 \ge 0, \ \left\|(x_0, x_1)\right\|_F = \begin{cases} x_0, & \text{if } x_1 \le x_0, \\ \dfrac{x_0^2 + x_1^2}{2x_1}, & \text{if } x_1 > x_0. \end{cases}$$

As a consequence, the set $S^+_{\|\cdot\|_F}$ corresponding to this norm is the set

$$S^+_{\|\cdot\|_F} = \left\{(1, x_1) : x_1 \in [0, 1]\right\} \cup \left\{ \left(x_0, 1 + \sqrt{1 - x_0^2}\right) : x_0 \in [0, 1) \right\}.$$

This set is represented in Figure 1.

Example 6.2 (Unit sphere for the Hüsler-Reiss norm). Let $\|\cdot\|_F$ be the bivariate Hüsler-Reiss norm with parameter σ^2, that is

$$\forall x, y > 0, \ \left\|(x, y)\right\|_F = x\Phi\left(\frac{\sigma}{2} + \frac{\log(x/y)}{\sigma}\right) + y\Phi\left(\frac{\sigma}{2} + \frac{\log(y/x)}{\sigma}\right).$$

Clearly $(1, 0)$ and $(0, 1)$ belong to $S^+_{\|\cdot\|_F}$. If $x, y > 0$ are such that $(x, y) \in S^+_{\|\cdot\|_F}$ then

$$\Phi\left(\frac{\sigma}{2} + \frac{\log(x/y)}{\sigma}\right) + \frac{y}{x}\Phi\left(\frac{\sigma}{2} + \frac{\log(y/x)}{\sigma}\right) = \frac{1}{x},$$

which implies, if $\lambda := y/x \in (0, \infty)$, that

$$x = \frac{1}{\Phi\left(\dfrac{\sigma}{2} - \dfrac{\log(\lambda)}{\sigma}\right) + \lambda\,\Phi\left(\dfrac{\sigma}{2} + \dfrac{\log(\lambda)}{\sigma}\right)},$$

$$\text{and } y = \frac{\lambda}{\Phi\left(\dfrac{\sigma}{2} - \dfrac{\log(\lambda)}{\sigma}\right) + \lambda\,\Phi\left(\dfrac{\sigma}{2} + \dfrac{\log(\lambda)}{\sigma}\right)}.$$

It is readily checked that conversely, any point of the form

$$\frac{1}{\Phi\left(\dfrac{\sigma}{2} - \dfrac{\log(\lambda)}{\sigma}\right) + \lambda\,\Phi\left(\dfrac{\sigma}{2} + \dfrac{\log(\lambda)}{\sigma}\right)}(1, \lambda), \ \text{ for } \lambda \in (0, \infty)$$

belongs to $S^+_{\|\cdot\|_F}$, so that we have a parametrization of $S^+_{\|\cdot\|_F}$ making it possible to represent this set. This is done in Figure 2 for various values of σ. One can observe in this Figure that, as should be apparent from the parametrization, the limit $\sigma \downarrow 0$ produces the part of the sphere of the sup-norm on $\mathbb{R}^2$ contained in the upper right quadrant, while the limit $\sigma \to \infty$ yields the segment $\{(x, 1 - x), \ 0 \le x \le 1\}$, corresponding to the sphere of the L^1−norm.

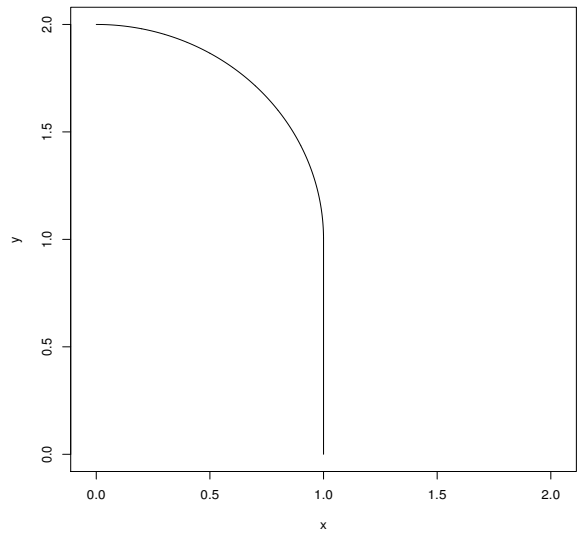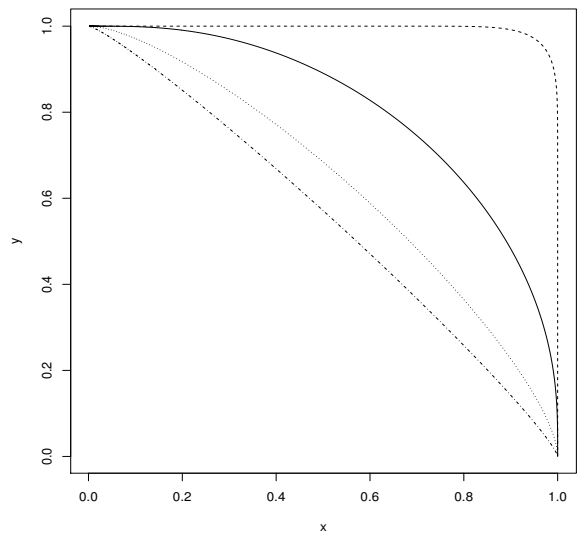**Figure 1:** The set $S^+_{\|\cdot\|_F}$ for the standard uniform distribution.

Figure 2: The set $S^+_{\|\cdot\|_F}$ for the bivariate Hüsler-Reiss F-norm. Dashed curve: $\sigma = 0.1$, solid curve: $\sigma = 1$, dotted curve: $\sigma = 2$, dashed-dotted curve: $\sigma = 3$.

Example 6.2 suggests that the convergence of F-norms, and thus convergence of the pertaining distributions in the Wasserstein metric, is at least informally linked to the convergence of their unit spheres. To make this intuition rigorous, we recall the definition of a *Hausdorff metric*. If $\|\cdot\|$ is an arbitrary norm on $\mathbb{R}^{d+1}$ and A, B are two subsets of $\mathbb{R}^{d+1}$, we let their $\|\cdot\|$-Hausdorff distance to be

$$d_{H,\|\cdot\|}(A, B) = \max \left\{ \sup_{\boldsymbol{y} \in B} \inf_{\boldsymbol{x} \in A} \|\boldsymbol{x} - \boldsymbol{y}\|, \ \sup_{\boldsymbol{x} \in A} \inf_{\boldsymbol{y} \in B} \|\boldsymbol{x} - \boldsymbol{y}\| \right\}.$$

Intuitively, two sets A and B are therefore close in the $\|\cdot\|$-Hausdorff metric if and only if each point in A (resp. B) is close, in terms of $\|\cdot\|$, to at least one point in B (resp. A). Such a distance may be infinite under no further assumptions on A and B, but is always finite if A and B are bounded. With this definition in mind, we have the following result.

Theorem 6.2. *Pointwise convergence of a sequence of F-norms $\|\cdot\|_{F_n}$ to an F-norm $\|\cdot\|_F$ on $\mathbb{R}^{d+1}$ is equivalent to convergence of the sequence of sets $S^+_{\|\cdot\|_{F_n}}$ to $S^+_{\|\cdot\|_F}$ in any Hausdorff metric $d_{H,\|\cdot\|}$ on $\mathbb{R}^{d+1}$.*

Our final result, relating convergence of distributions in the Wasserstein metric to convergence of unit spheres of F-norms in the Hausdorff metric, is now an immediate corollary of Theorems 3.1, 6.2 and the radial symmetry of F-norms.

Corollary 6.3. *If F_n, F are multivariate dfs on $\mathbb{R}^d$ with nonnegative, nonzero and integrable components, then the convergence of F_n to F in the Wasserstein metric is equivalent to the convergence of the unit sphere of $\|\cdot\|_{F_n}$ to the unit sphere of $\|\cdot\|_F$ in any Hausdorff metric $d_{H,\|\cdot\|}$ on $\mathbb{R}^{d+1}$.*

Proof. We start by noting that since all norms are equivalent on $\mathbb{R}^{d+1}$, it is sufficient to prove the theorem for the Hausdorff metric $d_{H,\|\cdot\|}$ induced by the norm $\|\cdot\|_F$.

Suppose that $\|\cdot\|_{F_n} \to \|\cdot\|_F$ pointwise. Then, by Theorem 3.1, we have $F_n \to F$ in the Wasserstein metric. Let $\boldsymbol{X}^{(n)}$, $\boldsymbol{X}$ have dfs F_n and F. This yields

$$\forall j \in \{1, \ldots, d\}, \ E\left(X_j^{(n)}\right) \to E\left(X_j\right) \ \text{ as } \ n \to \infty,$$

and thus, since $\boldsymbol{X}^{(n)}$, $\boldsymbol{X}$ satisfy $(\mathcal{H})$, there is $c > 0$ such that $E\left(X_j\right) \geq c$ and $E(X_j^{(n)}) \geq c$ for any n. Define then a weighted sup-norm $\|\cdot\|_{\infty,c}$ on $\mathbb{R}^{d+1}$ by

$$\left\|(x_0, x_1, \ldots, x_d)\right\|_{\infty,c} := \max(|x_0|, c|x_1|, \ldots, c|x_d|).$$

By Proposition 2.4, we obtain $\|\cdot\|_F \geq \|\cdot\|_{\infty,c}$ and $\|\cdot\|_{F_n} \geq \|\cdot\|_{\infty,c}$ for any n. Consequently, if B denotes the closed unit ball for the norm $\|\cdot\|_{\infty,c}$ and $\mathcal{B} := B \cap [0, \infty)^{d+1}$, then $\mathcal{B}$ contains $S^+_{\|\cdot\|_F}$ and the $S^+_{\|\cdot\|_{F_n}}$ for any n. In addition, by inequality (3) and since $\mathcal{B}$ is compact,

$$u_n := \sup_{\boldsymbol{x} \in \mathcal{B}} |\, \|\boldsymbol{x}\|_{F_n} - \|\boldsymbol{x}\|_F \,| \leq \sup_{\boldsymbol{x} \in \mathcal{B}} \|\boldsymbol{x}\|_\infty \cdot d_W(F_n, F) \to 0.$$

Assume from now on that n is so large that $u_n < 1$. Pick $\boldsymbol{x}$ in $S^+_{\|\cdot\|_F}$. Then since $\mathcal{B}$ contains $S^+_{\|\cdot\|_F}$, we have

$$|\, \|\boldsymbol{x}\|_{F_n} - 1 \,| = |\, \|\boldsymbol{x}\|_{F_n} - \|\boldsymbol{x}\|_F \,| \leq \sup_{\boldsymbol{x} \in \mathcal{B}} |\, \|\boldsymbol{x}\|_{F_n} - \|\boldsymbol{x}\|_F \,| = u_n.$$

This also entails $\|\boldsymbol{x}\|_{F_n} \geq 1 - u_n$. Note then that $\boldsymbol{x}/\|\boldsymbol{x}\|_{F_n} \in S^+_{\|\cdot\|_{F_n}}$ and thus

$$\left\| \boldsymbol{x} - \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_{F_n}} \right\|_F = \frac{|\, \|\boldsymbol{x}\|_{F_n} - 1 \,|}{\|\boldsymbol{x}\|_{F_n}} \leq \frac{u_n}{1 - u_n} =: \varepsilon_n. \tag{6}$$

If $\boldsymbol{x}$ in $S^+_{\|\cdot\|_{F_n}}$ we have, since $\mathcal{B}$ contains $S^+_{\|\cdot\|_{F_n}}$,

$$|\, \|\boldsymbol{x}\|_F - 1 \,| = |\, \|\boldsymbol{x}\|_{F_n} - \|\boldsymbol{x}\|_F \,| \leq \sup_{\boldsymbol{x} \in \mathcal{B}} |\, \|\boldsymbol{x}\|_{F_n} - \|\boldsymbol{x}\|_F \,| = u_n.$$

Write then $\boldsymbol{x}/\|\boldsymbol{x}\|_F \in S^+_{\|\cdot\|_F}$, which yields

$$\left\| \boldsymbol{x} - \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_F} \right\|_F = \big|\, \|\boldsymbol{x}\|_F - 1 \big| \le u_n \le \varepsilon_n. \tag{7}$$

From (6) and (7) it follows that

$$d_{H,\|\cdot\|_F}\left(S^+_{\|\cdot\|_{F_n}}, S^+_{\|\cdot\|_F} \right) \le \varepsilon_n \to 0,$$

showing the convergence of $S^+_{\|\cdot\|_{F_n}}$ to $S^+_{\|\cdot\|_F}$ in the Hausdorff metric $d_{H,\|\cdot\|_F}$.

Conversely, suppose that $S^+_{\|\cdot\|_{F_n}} \to S^+_{\|\cdot\|_F}$ in the Hausdorff metric $d_{H,\|\cdot\|_F}$. By radial symmetry and homogeneity of F-norms it is enough to prove the desired pointwise convergence of $\|\cdot\|_{F_n}$ to $\|\cdot\|_F$ on $S^+_{\|\cdot\|_F}$. Pick then $\boldsymbol{x} \in S^+_{\|\cdot\|_F}$. Note that $\boldsymbol{x}/\|\boldsymbol{x}\|_{F_n} \in S^+_{\|\cdot\|_{F_n}}$ and thus, by assumption, there is a sequence $(\boldsymbol{z}^{(n)}) \subset S^+_{\|\cdot\|_F}$ with

$$\left\| \boldsymbol{z}^{(n)} - \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_{F_n}} \right\|_F \to 0.$$

By the reverse triangle inequality, this entails

$$\left| 1 - \frac{1}{\|\boldsymbol{x}\|_{F_n}} \right| = \left| \left\| \boldsymbol{z}^{(n)} \right\|_F - \frac{\|\boldsymbol{x}\|_F}{\|\boldsymbol{x}\|_{F_n}} \right| \le \left\| \boldsymbol{z}^{(n)} - \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_{F_n}} \right\|_F \to 0.$$

This shows that $1/\|\boldsymbol{x}\|_{F_n} \to 1$ and thus $\|\boldsymbol{x}\|_{F_n} \to 1 = \|\boldsymbol{x}\|_F$ as required. $\qquad\square$

7 Conclusion

D-norms are tailor-made for multivariate extreme value theory as they turn out to provide an easily accessible common thread, in the sense that they do not require the knowledge of multivariate regular variation and of the associated topology background. Our paper introduces F-norms, which are an offspring of D-norms, to address the general framework of multivariate distributions rather than the max-stable distributions that are the focus of multivariate extreme value theory. While there is currently no competitor to the concept of D-norms in multivariate extreme value theory, there are of course various competitors to the concept of F-norms, such as the Fourier and Laplace transforms. As this paper shows, F-norms have their place in the probabilistic toolbox; while the pointwise convergence of Fourier and Laplace transforms is linked to weak convergence, the convergence of a pointwise sequence of F-norms translates into Wasserstein convergence of the underlying sequence of distributions. In this sense, F-norms behave like the max-characteristic function of [6], but their added norm structure offers an interesting geometric characterization of a distribution by a compact set, in a different way to existing alternatives such as the lift zonoid. As a corollary, the use of F-norms provides a nice interpretation of convergence in the Wasserstein metric by the means of convergence in Hausdorff metrics.

A promising aspect of F-norms is the good behavior of their random sample counterparts, as illustrated in Section 3, and the associated consequences this may have for statistical methodology. For instance, it is not difficult to imagine a goodness-of-fit test based on Theorem 3.3. The generalization of the results of Section 3 on the convergence of empirical F-norms to the case of stationary but dependent data is another interesting problem. Theorem 3.2 rests on the strong law of large numbers, which is known to be true for stationary and, say, mixing sequences under appropriate conditions, see for instance [14] in the context of ρ−mixing. Generalizing Theorem 3.3 would require the use of empirical processes techniques for dependent data, for which a good starting point is [3]. These lines of investigation are beyond the scope of the paper and will be part of future research on the topic of F-norms.

References

[1]   Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, Chichester.

[2]   Billingsley, P. (1999). *Convergence of Probability Measures*. Second edition. John Wiley & Sons, New York.

[3]   Dehling, H. and W. Philipp (2002). Empirical process techniques for dependent data. In H. Dehling, T. Mikosch, and M. Sørensen (Eds.), *Empirical Process Techniques for Dependent Data*, pp. 3–113. Birkhäuser, Boston MA.

[4]   Falk, M. (2019). *Multivariate Extreme Value Theory and D-Norms*. Springer, Cham.

[5]   Falk, M., J. Hüsler, and R.-D. Reiss (2011). *Laws of Small Numbers: Extremes and Rare Events*. Third edition. Birkhäuser, Basel.

[6]   Falk, M. and G. Stupfler (2017). An offspring of multivariate extreme value theory: The max-characteristic function. *J. Multivariate Anal. 154*, 85–95.

[7]   Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*. Second edition. Krieger, Malabar FL.

[8]   Gudendorf, G. and J. Segers (2010). Extreme-value copulas. In P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik (Eds.), *Copula Theory and Its Applications*, pp. 127–145. Springer, Berlin.

[9]   Koshevoy, G. and K. Mosler (1998). Lift zonoids, random convex hulls and the variability of random vectors. *Bernoulli 4*(3), 377–399.

[10] Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via KMT constructions. *Ann. Probab. 17*(1), 266–291.

[11] Mosler, K. (2002). *Multivariate Dispersion, Central Regions and Depth*. Springer, New York.

[12] Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press.

[13] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York.

[14] Shao, Q.-M. (1995). Maximal inequalities for partial sums of ρ-mixing sequences. *Ann. Probab. 23*(2), 948–965.

[15] Shorack, G. A. and J. A. Wellner (1986). *Empirical Processes with Applications to Statistics*. John Wiley & Sons, New York.

[16] Villani, C. (2009). *Optimal Transport*. Springer, Berlin.