The Metrics of Prokhorov and Ky Fan for Assessing Uncertainty in Inverse Problems

Andreas Hofinger*

July 21, 2006

Abstract

To assess the quality of solutions in stochastic inverse problems, a proper measure for the distance of random variables is essential.

The aim of this note is the comparison of the metrics of Ky Fan and Prokhorov with other concepts such as expected values, probability estimates and almost sure convergence.

In ill-posed problems one aims to find an appropriate solution $x^\dagger$ to an equation of the form

$$F(x) = y,$$

when the operator F is not continuously invertible. Therefore, the problems of interest are unstable; when only noisy data y^δ are available, special techniques (so called regularization methods) must be applied to obtain regularized solutions x_α^δ that are reasonable approximations to $x^\dagger$. To assess the quality of different regularization methods, in the theory of ill-posed problems convergence rate results, i. e., results of the form

$$\left\|x^\dagger - x_\alpha^\delta\right\| = \mathcal{O}\left(f(\left\|y - y^\delta\right\|)\right)$$

are an accepted quality criterion (see [12] for an introduction into this topic). So in a nutshell, in the deterministic theory of inverse problems the aim is to bound the distance between desired and regularized solution, in terms of the distance between exact and noisy data.

*Spezialforschungsbereich SFB F013/F1308, Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, A-4040 Linz, Austria email: andreas.hofinger@oeaw.ac.at

When the deterministic theory of inverse problems is to be extended to a stochastic setup, a question of utmost importance is how to measure distances of random variables, since now $x^\dagger$, x_α^δ, y, ... are replaced by their stochastic counterparts $x^\dagger(\omega)$, $x_\alpha^\delta(\omega)$, $y(\omega)$, ... In [18] an approach was presented that performs this extension using the metrics of Ky Fan and Prokhorov. In the following we collect some general results from [18] about these metrics.

The first section of this note introduces the metrics of Ky Fan and Prokhorov and describes general relations and differences between these metrics. In a second section we briefly compare the Ky Fan metric with other qualitative and quantitative concepts for measuring convergence. The final section gives a detailed quantitative comparison of convergence in expectation and in the Ky Fan metric.

1 The Metrics of Prokhorov and Ky Fan

Let us first introduce the metrics of Prokhorov and Ky Fan. As we will see later, these two are closely related, but while the latter works with random variables, the Prokhorov metric is concerned only with the underlying distributions.

1.1 The Prokhorov Metric

This metric does not directly work on the space of random variables, but on the underlying induced *distributions*. Suppose we are given two random variables $x(\omega)$ and $\tilde{x}(\omega) \in X$ for $\omega \in \Omega$ and probability space $(\Omega, \mathcal{A}, \mu)$. Via the measure μ on Ω we can define two corresponding measures μ_x and $\mu_{\tilde{x}}$ (the so-called *distributions* of x and $\tilde{x}$) on the space X: For a Borel set $B \subset X$ we define

$$\mu_x(B) := \mu(x^{-1}(B)) := \mu\{\omega \in \Omega \mid x(\omega) \in B\},$$

and

$$\mu_{\tilde{x}}(B) := \mu(\tilde{x}^{-1}(B)) := \mu\{\omega \in \Omega \mid \tilde{x}(\omega) \in B\}.$$

Instead of measuring the pointwise distance of $x(\omega)$ and $\tilde{x}(\omega)$ directly, we can use this lifting onto spaces of probability measures and compute the distance of the respective measures there, using an appropriate metric.

Distances between probability measures can be defined in numerous ways, see e. g. [17] for an overview. In [18] the aim was to develop a theory for stochastic inverse problems, that contains the deterministic one as a special case. For this sake a concept is needed that is applicable to point-measures, because such point-measures correspond to "constant random variables", which are essentially deterministic quantities. As the following remarks show, it is therefore important that the chosen metric metrizes the weak-star topology.

Remark 1.1. Consider an interval $I \subset \mathbb{R}$ and suppose that a sequence $(x_k)_{k=1}^{\infty}$, $x_k \in I$ converges to some $x \in I$. For which topology do the corresponding point-measures δ_{x_k} converge?

Every probability measure μ defines a continuous linear functional on the space of continuous functions $C(I)$ via

$$\mu(f) := \int_I f(x)\, d\mu(x)\,,$$

therefore every measure is an element of the dual space $C(I)^*$ of $C(I)$. Since $C(I)$ is not reflexive (cf. [28, 30]) there are at least 3 different possibilities for defining convergence of measures in $C(I)^*$: norm-, weak-, and weak-star-convergence. Let in the following δ_{x_k} and δ_x denote point-measures, associated with the points in the sequence $(x_k) \subset I \setminus \{x\}$ and their limit x.

- *Norm topology:* A norm on $C(I)^*$ can be defined via elements of the pre-dual space $C(I)$ as

  $$\|\mu_1 - \mu_2\| = \sup_{f \in C(I)} \frac{|\int f(x)\, d(\mu_1 - \mu_2)|}{\|f\|_{C(I)}}.$$

  For given k we can always find a continuous function $f_k(\cdot)$, $\|f_k\| \leq 1$, with $f_k(x) = 1$ and $f_k(x_k) = 0$ (e. g. a piecewise linear interpolant). Since $\int f_k(x)\, d(\delta_x - \delta_{x_k}) = f_k(x) - f_k(x_k)$, the distance between δ_{x_k} and δ_x in the norm remains constant and equal to 1; the point-measures δ_{x_k} do not converge to δ_x in the norm topology.

- *Weak topology:*[1] To investigate weak convergence we consider convergence of $\phi(\delta_{x_k})$ to $\phi(\delta_x)$ where $\phi \in C(I)^{**}$, i. e., ϕ is an element of the dual of $C(I)^*$. One particular functional on $C(I)^*$ is defined by the point evaluation

  $$\begin{aligned} \phi_x : C(I)^* &\to \mathbb{R} \\ \mu &\mapsto \phi_x(\mu) := \mu(\{x\})\,. \end{aligned}$$

[1] Note that the common notion of *weak topology* for probability distributions typically means the weak-star topology discussed below; we use the functional analytic terminology.

This functional is linear and has norm 1, as can be seen via the Riesz representation theorem ([28, 30]). With this choice we obtain for all $k \in \mathbb{N}$ that $\phi_x(\delta_{x_k}) = 0$, while at the same time $\phi_x(\delta_x) = 1$, thus δ_{x_k} does not converge weakly to δ_x.

- *Weak-star topology:* Here we measure convergence of δ_{x_k} to δ_x by applying these measures to continuous functions f. For any $f \in C(I) \subsetneq C(I)^{**}$ we have

$$|\delta_{x_k}(f) - \delta_x(f)| := |\int f(x)\, d(\delta_{x_k} - \delta_x)| = |f(x_k) - f(x)|\,.$$

  Since f is continuous we obtain that $\delta_{x_k} \to \delta_x$, whenever $x_k \to x$, so the point-measures δ_{x_k} do converge to δ_x in the weak-star topology.

Thus, the weak-star topology is weak enough such that $\delta_{x_k} \to \delta_x$ whenever $x_k \to x$.

A complementary question is considered in the following remark.

Remark 1.2. Given a probability space $(\Omega, \mathcal{A}, \mu)$, we denote the space of probability measures μ_x, which are actually distributions of random variables x, $x : \Omega \to X$, by $\mathcal{M}_{(\Omega,\mathcal{A},\mu)}(X)$). This space is not necessarily equal to the space of all probability measures on X (denoted by $\mathcal{M}(X)$), since not every probability measure needs to be the distribution of a random variable on $(\Omega, \mathcal{A}, \mu)$.

But what can be said about a probability measure that is, in some sense, the limit of distributions of random variables, will it also be the distribution of a random variable? I.e., if a sequence μ_{x_n} converges to some $\tilde{\mu} \in \mathcal{M}(X)$, does there exist a random variable $\tilde{x}$, $\tilde{x} : \Omega \to X$, such that $\tilde{\mu} = \mu_{\tilde{x}}$? For the case of convergence in the weak-star topology, this question is answered affirmatively in [14].

Thus, the weak-star topology is strong enough such that $\mathcal{M}_{(\Omega,\mathcal{A},\mu)}(X)$ is a sequentially closed subset of $\mathcal{M}(X)$.

Although many different distances on probability spaces are available, only few of them metrize the weak-star topology. Among them are the Prokhorov metric and the Wasserstein or bounded Lipschitz metric. The result of Strassen (see Theorem 1.6) gives connections between the distance of the measures on the probability space measured in the Prokhorov metric, and the distance of corresponding random variables measured in the Ky Fan metric, and is the reason why we finally pick this metric.

Definition 1.3 (Prokhorov metric)**.** *The distance of two measures* μ_1, μ_2 *in the Prokhorov metric is defined as (see, e. g., [5, 9, 19, 22])*

$$\rho_{\mathrm{P}}(\mu_1,\, \mu_2) := \inf\{\varepsilon > 0 \mid \mu_1(B) \le \mu_2(B^\varepsilon) + \varepsilon,\, \forall\, \textit{Borel sets}\, B \subset \Omega\}\,. \quad (1)$$

Here $B^\varepsilon = \{x \mid d(x, B) < \varepsilon\}$*, where* $d(x, B)$ *is the distance of* x *to* B*, i. e.,* $d(x, B) = \inf_{z\in B} \|x - z\|$.

The use of B^ε in the definition above is essential: The Prokhorov distance of two measures μ_1 and μ_2 is small, when the probability of similar events (B^ε) is similar up to a small quantity ("$+\varepsilon$"). In contrast, the *total variation distance* (cf. [17]) measures if the probability of *the same* event (B) is similar ("$+\varepsilon$"). Consequently, when x_k, $x_k \neq x$, converges to x the corresponding point-measures δ_{x_k} converge to δ_x in the Prokhorov metric, but they do not in the total variation distance.

Having defined a metric on the space of probability measures, an interesting question is the following: Consider an operator $F : X \to Y$, and a sequence of random variables (x_k), with $x_k \to x$, where convergence is measured in the Prokhorov metric. Under which continuity assumptions on F does the random variable $F(x_k)$ converge to $F(x)$ (again in the Prokhorov metric), and can this convergence be quantified? This question is answered in [15, 29]. As demonstrated in [13, 15] the answer to this question is also relevant for regularization theory of stochastic inverse problems, because it can be used to derive convergence rate results.

1.2 The Ky Fan Metric

While the Prokhorov metric works with distributions, the Ky Fan metric uses random variables to define distances. This metric is defined as follows.

Definition 1.4 (Ky Fan metric)**.** *The distance of two random variables* ξ_1*,* ξ_2 *in the Ky Fan metric is defined as ([16], also [9])*

$$\rho_{\mathrm{K}}(\xi_1,\, \xi_2) := \inf\,\{\varepsilon > 0 \mid \mu\{\omega \in \Omega \mid d(\xi_1(\omega), \xi_2(\omega)) > \varepsilon\} < \varepsilon\}\,. \qquad (2)$$

Convergence in the Ky Fan metric is a quantitative version of *convergence in probability* (see below).

Let us first give a short interpretation of the Ky Fan distance. If ξ_1 and ξ_2 have distance $\rho_{\mathrm{K}}(\xi_1,\, \xi_2) \le \varepsilon$, this implies that

- with high probability (namely $1 - \varepsilon$) the realizations of the random variables have distance $d(\xi_1(\omega), \xi_2(\omega)) \le \varepsilon$,

- with low probability (at most ε), the distance between $\xi_1(\omega)$ and $\xi_2(\omega)$ may be larger than ε.

In particular, the second point is of interest: The Ky Fan distance between ξ_1 and ξ_2 may be small, although on a set of positive probability the distance of $\xi_1(\omega)$ and $\xi_2(\omega)$ might be arbitrarily large.

In contrast, the expected value is influenced by all events that have positive probability. In particular for non-normally distributed random variables this can make a significant difference. For instance in section 3 we construct an example for which $\rho_{\kappa}(\xi_1, \xi_2) \to 0$ while $\mathbb{E}\left(\|\xi_1 - \xi_2\|^2\right)$ remains constant or even tends to infinity. This is also relevant for convergence rates for stochastic inverse problems, as a numerical example in [18, Ch. 5] shows:

> The quality of the solutions there is measured by some parameter s, where $s = 0$ corresponds to a deterministic problem with well-behaved solution; with growing s the probability of "bad" solutions increases. While now the Ky Fan metric gives
>
> $$\rho_{\kappa}(x^{\delta}_{k_*(\delta)}, x^{\dagger})_{H^1(I)} = \mathcal{O}\left(\delta^{\frac{2\nu}{2\nu+1+s}}\right),$$
>
> i. e., a convergence rate that slows down gradually for increasing s, the rate in expectation is given as
>
> $$\mathbb{E}\left(\left\|x^{\dagger} - x^{\delta}_{k_*}\right\|^2_{H^1(I)}\right)^{1/2} \leq \begin{cases} \mathcal{O}(\delta^{\frac{2\nu}{2\nu+1}}) & 0 < s < \nu + \frac{1}{2} \\ \infty & \nu + \frac{1}{2} \leq s \,. \end{cases}$$
>
> i. e., it remains constant for a while, and switches to non-convergence suddenly.[2]

So on the one hand measuring convergence in expectation, one is restricted to a smaller class of random variables. On the other hand, the expected value does not give information about particular realizations, while the Ky Fan distance gives a good bound with probability $1 - \varepsilon$.

1.3 Relations between Prokhorov and Ky Fan Metric

Let us in the following consider some connections and differences between the Ky Fan metric and the Prokhorov metric.

In case we are interested in the distance of a random variable to a point, or respectively, the distance of a distribution to a point-measure, it turns out that the Ky Fan distance and the Prokhorov distance are equal.

[2] Of course the involved coefficients in the $\mathcal{O}(\cdot)$-notation do not remain constant. But since these constants are in practice unavailable, the focus in inverse-problems theory is on the appearing exponents in the convergence rate (see [12]).

Proposition 1.5. *Let ξ_1, ξ_2 be two random variables with distributions μ_1, μ_2. Let one of the random variables be constant. Then*

$$\rho_{\text{K}}(\xi_1,\, \xi_2) = \rho_{\text{P}}(\mu_1,\, \mu_2)\,.$$

Proof. It suffices to show that $\rho_{\text{K}}(\xi_1,\, \xi_2) \le \rho_{\text{P}}(\mu_1,\, \mu_2)$ (see Proposition 1.7 below for the converse estimate).

Suppose that ξ_1 is constant, i. e., $\xi_1(\omega) = x_1$ for almost all $\omega \in \Omega$. According to Definition 1.3 we have for arbitrary Borel-sets B and $\varepsilon > \rho_{\text{P}}(\mu_1,\, \mu_2)$

$$\mu_1(B) < \mu_2(B^\varepsilon) + \varepsilon\,.$$

in particular this also holds for the set $B = \{x_1\}$. For this choice we obtain

$$\begin{aligned} 1 - \varepsilon &< \mu_2(B^\varepsilon) \\ &= \mu\{\omega \in \Omega \mid \xi_2(\omega) \in B^\varepsilon\} \\ &= \mu\{\omega \in \Omega \mid d(\xi_2(\omega), x_1) \le \varepsilon\} \end{aligned}$$

Via Definition 1.4 this implies $\rho_{\text{K}}(\xi_1,\, \xi_2) \le \varepsilon$. Taking the infimum with respect to $\varepsilon > \rho_{\text{P}}(\mu_1,\, \mu_2)$ concludes the proof. □

If we are interested in the distance of two genuine random variables, more effort is necessary to connect the Prokhorov metric and the Ky Fan metric. The following theorem, originally obtained by Strassen [27] and extended by Dudley [8], is an important tool for this task. (The proof of the following result can also be found in [9, ch. 11.6].)

Theorem 1.6 (Strassen)**.** *Let $(\mathcal{X}, d)$ be a separable metric space and $\mathcal{M}(\mathcal{X})$ be the set of Borel probability measures on $\mathcal{X}$, μ_1, μ_2 be elements of this space. Let $\alpha \ge 0$ and $\beta \ge 0$. Then the following statements are equivalent*

(i). $\mu_1(B) \le \mu_2(B^{\alpha]}) + \beta$ *for all Borel sets B*

(ii). For any $\varepsilon > 0$ there exists $\mu_{1/2} \in \mathcal{M}(\mathcal{X} \times \mathcal{X})$ with $\mu_1 = \mu_{1/2} \circ \pi_1^{-1}$, $\mu_2 = \mu_{1/2} \circ \pi_2^{-1}$ and $\mu_{1/2}(d(x_1, x_2) > \alpha + \varepsilon) \le \beta + \varepsilon$. Here π_1, π_2 are the natural projections from $\mathcal{X} \times \mathcal{X}$ onto $\mathcal{X}$, i. e., $\pi_1(x_1, x_2) = x_1$, $\pi_2(x_1, x_2) = x_2$.

Here $B^{\alpha]}$ denotes the closure of B^α. If furthermore $\mathcal{X}$ is complete, this equivalence also holds for the case $\varepsilon = 0$.

This theorem immediately gives the following relation between the Prokhorov metric and the Ky Fan metric [9, 25, 27].

Proposition 1.7. *Let the assumptions of Theorem 1.6 be satisfied, and $\mathcal{X}$ be complete. Then with the notation of Theorem 1.6 the following statements are equivalent*

(i). The Prokhorov distance of two measures μ_1, μ_2, satisfies $\rho_{\mathrm{P}}(\mu_1, \mu_2) \leq \varepsilon$.

(ii). There exists random variables ξ_1 and ξ_2 such that μ_1 and μ_2 are the distributions of ξ_1 and ξ_2, respectively, and $\rho_{\mathrm{K}}(\xi_1, \xi_2) \leq \varepsilon$.

Thus, we always have $\rho_{\mathrm{P}}(\mu_1, \mu_2) \leq \rho_{\mathrm{K}}(\xi_1, \xi_2)$; for given μ_1, μ_2 we can find ξ_1, ξ_2 and a joint distribution $\mu_{1/2} \in \mathcal{M}(\mathcal{X} \times \mathcal{X})$ such that equality holds.

Proof. Let $\rho_{\mathrm{P}}(\mu_1, \mu_2) = \varepsilon$. Then by Definition 1.3 (note that the set on the right is closed)

$$\mu_1(B) \leq \mu_2(B^{\varepsilon]}) + \varepsilon\,.$$

The theorem of Strassen now guarantees existence of $\mu_{1/2} \in \mathcal{M}(\mathcal{X} \times \mathcal{X})$ with the property

$$\mu_{1/2}(d(x_1, x_2) > \varepsilon) \leq \varepsilon$$

so $\rho_{\mathrm{K}}(\pi_1, \pi_2) \leq \rho_{\mathrm{P}}(\mu_1, \mu_2)$, where the projections π_1 and π_2 are seen as random variables from $\mathcal{X} \times \mathcal{X}$ to $\mathcal{X}$. The distributions of π_1 and π_2 are μ_1 and μ_2, respectively.

For the converse implication, suppose $\rho_{\mathrm{K}}(\xi_1, \xi_2) = \varepsilon$, and choose $\delta > 0$. Via Strassen's theorem this implies

$$\mu_1(B) \leq \mu_2(B^{\varepsilon+\delta]}) + \varepsilon + \delta\,.$$

To obtain an open set on the right hand side, we can estimate

$$\mu_1(B) \leq \mu_2(B^{\varepsilon+2\delta}) + \varepsilon + 2\delta\,.$$

This holds for all Borel-sets B, and therefore $\rho_{\mathrm{P}}(\mu_1, \mu_2) \leq \rho_{\mathrm{K}}(\xi_1, \xi_2) + 2\delta$. The estimate follows by building the infimum with respect to $\delta > 0$. □

Since the Prokhorov distance of two measures μ_1 and μ_2 is so closely related to the Ky Fan distance of corresponding random variables ξ_1 and ξ_2, the appearance of many convergence results remains unchanged when the Prokhorov metric is replaced by the Ky Fan metric. For instance the structure of the lifting result presented in [13, Thm. 2.1] remains unchanged when $\rho_{\mathrm{P}}(\cdot, \cdot)$ is replaced by $\rho_{\mathrm{K}}(\cdot, \cdot)$ (see [18, Thm. 1.15]).

Let us now turn to *differences* between the Prokhorov metric and the Ky Fan metric.

As already noted, the Ky Fan distance gives information about concrete realizations of random variables, while the Prokhorov distance only works with the underlying distributions. To underline this difference, consider tossing a fair coin, where we denote the outcome “heads” by $+1$ and “tails” by -1. Now suppose that (maybe due to an assembly error) our measurement device always observes the opposite outcome, i. e., it measures “heads” when there is “tails” and vice versa. What is the measurement error for this experiment?

- For the Ky Fan metric the error clearly is equal to 1: In 100% of the cases we measure $+1$ instead of -1 and vice versa. So with probability 1 the measurement error is larger than 1 (cf. Definition 1.4).

- In contrast, for the Prokhorov metric the error is 0, i. e., the observation seems to be noise free(!): In 50% of the cases we observe the outcome $+1$, in 50% of the cases we observe the outcome -1. Although each single measurement is wrong, in total we find that the experiment describes a binomial distribution with parameter $\frac{1}{2}$, which is the correct observation.

This example indicates that results in the Ky Fan metric will typically be suitable when the focus is on random variables. In particular, an estimate in the Ky Fan metric gives a *confidence interval* for concrete realizations; if $\rho_{\text{K}}(\xi_1, \xi_2) \leq \varepsilon$, with probability $1 - \varepsilon$ the realizations $\xi_1(\omega)$ and $\xi_2(\omega)$ have distance less than ε.

In contrast, convergence results in the Prokhorov metric are of interest, when the observed data are a distribution. Consider a biological system where the growth behavior of cells is to be analyzed (cf. [1]). While it is possible to determine, e. g., the distribution of the cell sizes within the system (maybe even in a time-dependent way), there may be no means to track the behavior of only a single cell. Clearly, it is not possible to use this input data to determine parameters of individual cells; so one cannot ask for more than convergence of the reconstructed distribution to the correct value. The Prokhorov metric is the right tool to measure convergence in such a situation.

1.4 Further Properties of the Prokhorov metric

In this section we mention some additional properties of the Prokhorov metric. Suppose in the following that $\mathcal{X}$ is a vector space equipped with a metric, and $\mathcal{M}(\mathcal{X})$ is the set of probability measures on this space. Some strong relations hold between the space $\mathcal{X}$, and $\mathcal{M}(\mathcal{X})$ equipped with the metric $\rho_{\text{P}}(\cdot, \cdot)$; these results were obtained by Prokhorov in [23] (cf. [22]).

- $\mathcal{M}(\mathcal{X})$ can be metrized as a separable metric space if and only if $\mathcal{X}$ is itself a separable metric space.

- $\mathcal{M}(\mathcal{X})$ is a Polish space (i. e., separable, metric and *complete*) if and only if $\mathcal{X}$ is Polish.

- $\mathcal{M}(\mathcal{X})$ is a *compact* metric space if and only if $\mathcal{X}$ is compact metric.

The last fact was used in [1] to show that a certain approximation scheme leads to convergence in the Prokhorov metric: In the considered case an ODE in a biological system is influenced by some real valued random variable $C(\omega)$, each individual carrying a different realization of C. The goal in [1] was now to recover the distribution of $C(\omega)$ via a time-dependent measurement of the evolution of the system. Since in [1], $C(\omega)$ was confined to some closed interval, we obtain from the above statement that also the corresponding distributions come from a compact space, and thus, have convergent subsequences in the Prokhorov metric. If the desired quantity was no real or (finite dimensional) vector-valued random variable C, but a random function this approach would fail, since closed bounded sets in infinite dimensional spaces need not be compact.

2 Comparison with Other Concepts

In the following we consider various other approaches that are common when describing convergence of random variables. This comparison is split into qualitative and quantitative concepts. As qualitative ones we choose the two common concepts *convergence almost surely* and *convergence in probability*; as quantitative ones we choose such concepts that have been used in the theory of stochastic inverse problems (see e. g., [6, 26]).

2.1 Qualitative Concepts

Convergence almost surely: A sequence x_k converges to x almost surely (a. s.), when for almost all ω

$$\|x_k(\omega) - x(\omega)\| \to 0\,, \tag{3}$$

i. e., except on a null set, $x_k(\omega)$ converges point-wise to $x(\omega)$. Almost sure convergence implies convergence in the Ky Fan metric; the converse is true for subsequences (Propositions 2.1 and 2.2 below).

Convergence in Probability: A sequence x_k converges to x in probability, when for all $\varepsilon > 0$

$$\mu\{\omega \in \Omega \mid \|x_k(\omega) - x(\omega)\| > \varepsilon\} \to 0\,.$$

Comparing this definition with (2), it can be seen easily, that this type of convergence is equivalent to convergence in the Ky Fan metric (see also [16]).

Relations between almost sure convergence and convergence in probability are well-known (see e. g., [2, 3, 7, 9, 11]). Because convergence in the Ky Fan metric is a quantitative version of convergence in probability analogous results hold for the Ky Fan metric as well. In the following we discuss these relations between almost sure convergence and convergence in the Ky Fan metric. The first theorem follows immediately from the analogous result for convergence in probability.

Proposition 2.1. *Let $x_k \to x$ almost surely. Then $\rho_{\kappa}(x_k, x) \to 0$.*

The converse result is not true. Nevertheless, it is well-known that convergence in probability implies almost sure convergence at least of *subsequences* (see e. g., [9, 11]). The next proposition gives a quantitative version of this statement for the Ky Fan metric[3].

Proposition 2.2. *Let x_k converge to x in the Ky Fan metric. Then for any $\eta > 0$ and $\varepsilon > 0$ there exist Ω_ε, $\mu(\Omega_\varepsilon) \geq 1 - \varepsilon$, and a subsequence x_{k_j} with*

$$\left\|x_{k_j}(\omega) - x(\omega)\right\| \leq (1+\eta)\rho_{\kappa}(x_{k_j}, x) \quad \text{for all } \omega \in \Omega_\varepsilon.$$

Furthermore there exists a subsequence that converges to x almost surely.

Proof. Set $\delta_k := (1+\eta)\rho_{\kappa}(x_k, x)$. By Definition 1.4, for given δ_k, there exists a set Ω_{δ_k} with

$$\mu\left(\Omega_{\delta_k}\right) \geq 1 - \delta_k, \quad \text{and} \quad \omega \in \Omega_{\delta_k} \implies \|x(\omega) - x_k(\omega)\| \leq \delta_k\,.$$

In general we cannot deduce convergence of $x_k(\omega)$ to $x(\omega)$ for $\omega \in \Omega_{\delta_k}$ and $\delta_k \to 0$, since the sets Ω_{δ_k} may have empty intersection for $\delta_k \to 0$. Thus we need the following construction. For arbitrary $\varepsilon > 0$ and $\delta_k \to 0$ we pick a subsequence (δ_{k^j}) with $\sum_{j=1}^{\infty} \delta_{k^j} \leq \varepsilon$, and introduce the set

$$\Omega_\varepsilon := \bigcap_{j=1}^{\infty} \Omega_{\delta_{k^j}}\,,$$

[3] Presumably, the following result has not been explicitly derived before, since only little attention was paid to the Ky Fan metric in the past.

which is a subset of every $\Omega_{\delta_{k^j}}$. This set has measure $\mu(\Omega_\varepsilon) \geq 1 - \varepsilon$ since

$$\mu\left(\bigcap_{j=1}^{\infty} \Omega_{\delta_{k^j}}\right) = \mu\left(\Omega \setminus \bigcup_{j=1}^{\infty} \left(\Omega \setminus \Omega_{\delta_{k^j}}\right)\right) = 1 - \mu\left(\bigcup_{j=1}^{\infty} \left(\Omega \setminus \Omega_{\delta_{k^j}}\right)\right)$$
$$\geq 1 - \sum_{j=1}^{\infty} \mu\left(\Omega \setminus \Omega_{\delta_{k^j}}\right) \geq 1 - \varepsilon$$

Since Ω_ε is a subset of every $\Omega_{\delta_{k^j}}$ we have

$$\forall \omega \in \Omega_\varepsilon \subseteq \Omega_{\delta_{k^j}} : \quad \left\| x(\omega) - x_{k_j}(\omega) \right\| \leq \delta_{k^j} \,,$$

which proves the first statement.

For the second statement, consider the set N on which x_{k_j} does not converge to x. This set is given as

$$\begin{aligned} N &= \left\{\omega \mid \exists \tilde{\varepsilon} > 0 \, \forall j_0 \in \mathbb{N} \, \exists j \geq j_0 : \, \left\| x_{k_j}(\omega) - x(\omega) \right\| \geq \tilde{\varepsilon} \right\} \\ &= 1 - \left\{\omega \mid \forall \tilde{\varepsilon} > 0 \, \exists j_0 \in \mathbb{N} \, \forall j \geq j_0 : \, \left\| x_{k_j}(\omega) - x(\omega) \right\| < \tilde{\varepsilon} \right\} \,. \end{aligned}$$

Similarly to Ω_ε above, we define the set Ω_{j_0} as

$$\begin{aligned} \Omega_{j_0} := \bigcap_{j \geq j_0} \Omega_{\delta_{k^j}} &= \left\{\omega \mid \forall j \geq j_0 : \, \left\| x_{k_j}(\omega) - x(\omega) \right\| < \delta_{k^j} \right\} \\ &\subseteq \left\{\omega \mid \forall \tilde{\varepsilon} > 0 \, \exists j_0 \in \mathbb{N} \, \forall j \geq j_0 : \, \left\| x_{k_j}(\omega) - x(\omega) \right\| < \tilde{\varepsilon} \right\}. \end{aligned}$$

Since the sequence x_{k_j} is uniformly convergent to x on the set Ω_{j_0}, we can estimate

$$\begin{aligned} \mu(N) &= 1 - \mu\left(\left\{\omega \mid \forall \tilde{\varepsilon} > 0 \, \exists j_0 \in \mathbb{N} \, \forall j \geq j_0 : \, \left\| x_{k_j}(\omega) - x(\omega) \right\| < \tilde{\varepsilon} \right\}\right) \\ &\leq 1 - \mu\left(\left\{\omega \mid \forall j \geq j_0 : \, \left\| x_{k_j}(\omega) - x(\omega) \right\| < \delta_{k^j} \right\}\right) \\ &= 1 - \mu\left(\bigcap_{j \geq j_0} \Omega_{\delta_{k^j}}\right) \\ &= \mu\left(\bigcup_{j \geq j_0} \Omega \setminus \Omega_{\delta_{k^j}}\right) \leq \sum_{j \geq j_0} \delta_{k^j} \,. \end{aligned}$$

Since $\sum \delta_{k^j} \leq \varepsilon$, and this sum is absolutely convergent, we obtain

$$\sum_{j \geq j_0} \delta_{k^j} \to 0 \quad \text{as} \quad j_0 \to \infty \,.$$

Hence N is a null set; x_{k_j} converges to x almost surely. $\square$

2.2 Quantitative Concepts

Now we turn to quantitative concepts to measure convergence of random variables. Two common concepts are expected values, and probability estimates.

Convergence in Expectation: Here we look for bounds on the expected value of the distance, defined as

$$\mathbb{E}\left(\|x_k - x\|^2\right) := \int_\Omega \|x_k(\omega) - x(\omega)\|^2 \, d\mu(\omega)\,. \tag{4}$$

For the case $X = L_2(I)$, we find that $\mathbb{E}\left(\|\cdot\|^2\right)$ is a (weighted) norm on the product space $L_2(\Omega \times I)$.

Probability Estimates: Similarly as for the Ky Fan metric, the space Ω is split into parts with high and low probability, the resulting estimates have the form[4]

$$\mathbb{P}\left(\|x_k - x\| \le \varepsilon_1(p)\right) > 1 - \varepsilon_2(p)\,. \tag{5}$$

Here $\varepsilon_1(p)$ and $\varepsilon_2(p)$ are functions of one or more parameters p. Typically, these functions are continuous.

The concept of expectation can be a too restrictive notion of convergence. A first indication is the fact that almost sure convergence as in (3) does not guarantee that also $\mathbb{E}\left(\|x_k - x\|^2\right)$ tends to 0; even worse, we could have that $\mathbb{E}\left(\|x_k - x\|^2\right) = \infty$ and remains unbounded, no matter how large k is chosen (cf. the examples in section 3).

Therefore the second error measure can be considered more natural. Assuming that x_k converges to x almost surely, we have for any fixed ε_1 that $\mathbb{P}\left(\|x_k - x\| \le \varepsilon_1\right)$ tends to 1. Vice versa, for fixed ε_2 and arbitrary small $\varepsilon_1 > 0$ we can find $k \in \mathbb{N}$ with $\mathbb{P}\left(\|x_k - x\| \le \varepsilon_1\right) > 1 - \varepsilon_2$. A drawback of this concept is that *two* parameters are necessary to describe the distance; estimates of the form (5) do not form a metric.

But fortunately, via Definition 1.4 these can be translated into estimates in the Ky Fan metric, and via Proposition 1.7 also into the Prokhorov metric. To do this, one has to solve the equation

$$\varepsilon_1(p) = \varepsilon_2(p)$$

---

[4] Depending on the context we will sometimes use the notion of probabilities instead of measures, but of course these are only two different words for the same meaning and $\mathbb{P}(B) \equiv \mu(B)$.

for p. The resulting solution gives some $\varepsilon(p) = \varepsilon_1(p) = \varepsilon_2(p)$ with $\rho_\kappa(x_k, x) \le \varepsilon(p)$. The task of solving this equation is often non-trivial, but as its outcome we obtain a distance between x_k and x in terms of a metric, which allows us to investigate the speed of convergence.

Thus, concluding we will find that the Ky Fan metric gives

- a concept that is better suited for treating stochastic problems than the expected value, especially when the distributions of the appearing variables may have 'fat tails', i. e., when there is a higher probability for large values to occur than for a normally distributed random variable. In particular, the Ky Fan distance is always finite, whereas the expected value (4) can be unbounded; pointwise convergence implies convergence in the Ky Fan metric (cf. e. g. the constructions in the next section).

- a framework that translates estimates as in (5) to an interpretable setup. This allows comparison of new stochastic results with the classical deterministic ones.

- a setting that contains the deterministic results as a special case.

Let us in the following make a more detailed comparison between the Ky Fan metric and the expected value.

3 Expectation vs. Ky Fan metric

3.1 Non-Convergence of Expectation

In the following we construct a sequence x_k that converges to 0 in the Ky Fan metric, but does not converge in expectation. For this construction we define the random variable $x(\omega)$ as

$$x(\omega) = \frac{C(\alpha)}{\omega^\alpha}, \tag{6}$$

where ω is uniformly distributed on $\Omega = [0, 1]$. Observe that $x(\omega)$ is unbounded on the interval $[0, 1]$ for $\alpha > 0$; nevertheless, clearly for $C(\alpha) \to 0$ also $x(\omega)$ tends to 0 on $(0, 1]$ point-wise.

We now compute the expected value of the distance of $x(\omega)$ to 0. In the first case we look for the expectation of the absolute value of the error and obtain

$$\mathbb{E}\left(|x-0|\right) = \int_\Omega \frac{C(\alpha)}{\omega^\alpha}\, d\omega = C(\alpha)\int_\Omega \omega^{-\alpha}\, d\omega$$
$$= \begin{cases} \frac{C(\alpha)}{1-\alpha} & \alpha < 1 \\ \infty & \alpha \geq 1\,. \end{cases}$$

For the second case we investigate the square of the absolute value

$$\mathbb{E}\left(|x-0|^2\right) = \int_\Omega \frac{C(\alpha)^2}{\omega^{2\alpha}}\, d\omega = C(\alpha)^2\int_\Omega \omega^{-2\alpha}\, d\omega$$
$$= \begin{cases} \frac{C(\alpha)^2}{1-2\alpha} & \alpha < \frac{1}{2} \\ \infty & \alpha \geq \frac{1}{2}\,. \end{cases}$$

Observe that, for α sufficiently large, in both cases the expected value may be infinite, although $x(\omega)$ tends to 0 pointwise.

Now we compute the Ky Fan distance, therefore we have to find the smallest t such that

$$\mathbb{P}\left(|x(\omega)-0| > t\right) < t.$$

Inserting (6) we have for the term on the left

$$\mathbb{P}\left(\frac{C(\alpha)}{\omega^\alpha} > t\right) = \mathbb{P}\left(\frac{1}{\omega} > \left(\frac{t}{C(\alpha)}\right)^{1/\alpha}\right) = \left(\frac{C(\alpha)}{t}\right)^{1/\alpha},$$

which further leads to the relation

$$\left(\frac{C(\alpha)}{t}\right)^{1/\alpha} = t\,.$$

Solving this equation for t we obtain for the Ky Fan distance the expression

$$\rho_\kappa(x,\,0) = C(\alpha)^{\frac{1}{1+\alpha}}.$$

Now define a sequence of random variables $x_k(\omega) = \frac{C(\alpha_k)}{\omega^{\alpha_k}}$ with some $\alpha_k \nearrow 1$ and $C(\alpha_k) := 1-\alpha_k$. Then we obtain that in the Ky Fan metric this sequence tends to 0, indeed

$$\rho_\kappa(x_k,\,0) = (1-\alpha_k)^{\frac{1}{1+\alpha_k}} \leq (1-\alpha_k)^{\frac{1}{2}} \to 0\,. \tag{7}$$

So the random variable x_k converges to 0 in the Ky Fan metric, nevertheless the expectation does not tend to 0, it even remains constant

$$\mathbb{E}\left(|x_k-0|\right) = \frac{C(\alpha_k)}{1-\alpha_k} = 1 \not\to 0\,. \tag{8}$$

The choice $\alpha_k \nearrow \frac{1}{2}$ and $C(\alpha_k) := 1 - 2\alpha_k$ yields the analogous result for $\mathbb{E}\left(|x_k - 0|^2\right)$. Observe that for the latter choice $\mathbb{E}\left(|x_k - 0|^2\right)$ remains constant while $\mathbb{E}\left(|x_k - 0|\right)$ tends to 0. Furthermore observe that $C(\alpha_k) := (1-\alpha_k)^{1/2}$ even leads to divergence in (8), while maintaining convergence in (7).

Clearly, the presented results arise from the unboundedness of the involved random variables. If the distance of the random variables ξ_1 and ξ_2 is bounded a-priori, we obtain from (2) (with arbitrary $\varepsilon > \rho_\kappa(\xi_1, \xi_2)$)

$$\begin{aligned}\mathbb{E}\left(\|\xi_1 - \xi_2\|^2\right) &:= \int_\Omega \|\xi_1(\omega) - \xi_2(\omega)\|^2 \, d\mu(\omega) \\ &\leq (1-\varepsilon)\varepsilon^2 + \varepsilon \sup_{\omega\in\Omega} \|\xi_1(\omega) - \xi_2(\omega)\|^2 \leq C\varepsilon \, .\end{aligned}$$

So for bounded random variables, convergence in the Ky Fan metric implies convergence of the expected value.

3.2 Convergence via Markov's Inequality

We have seen that a sequence may converge in the Ky Fan metric, without the need to converge in expectation. On the other hand, comparing the representation (4) with (2) it seems clear that whenever we observe convergence in expectation, also the Ky Fan distance will tend to 0. But can this statement be quantified?

For "constant" random variables, i. e., the deterministic case, it turns out that the convergence rate in the Ky Fan metric is the same as the rate observed for the expected value. Furthermore, for the examples in this note we find that the convergence rate in the Ky Fan metric is always slower, than the rate observed for the expected value (if the expected value converges).[5]

A tool to quantify these statements is Markov's inequality ([21], cf. [10]).

Theorem 3.1 (Markov)**.** *For any non-negative random variable X and $c > 0$ we have the estimate*

$$\mathbb{P}\left(X \geq c\right) \leq \frac{\mathbb{E}\left(X\right)}{c} \, . \tag{9}$$

From this theorem we obtain, by setting $X = \|x\|^s$ with $s > 0$

$$\mathbb{P}\left(\|x\| \geq c\right) = \mathbb{P}\left(\|x\|^s \geq c^s\right) \leq \frac{\mathbb{E}\left(\|x\|^s\right)}{c^s} \, . \tag{10}$$

Using this inequality, we can determine bounds on the Ky Fan distance in terms of the expected value as follows.

[5]see also the results in [18, Ch. 5.6]

Theorem 3.2. *Let μ_1 and μ_2 be the distributions of two random variables ξ_1 and ξ_2. The Ky Fan and the Prokhorov distance of the distributions can be bounded via the expected value of the distance of the random variables as follows*

$$\rho_{\mathrm{P}}(\mu_1,\, \mu_2) \le \rho_{\mathrm{K}}(\xi_1,\, \xi_2) \le \sqrt{\mathbb{E}\left(\|\xi_1 - \xi_2\|\right)}\,.$$

In particular, for arbitrary $s > 0$ (with possibly infinite right hand side)

$$\rho_{\mathrm{P}}(\mu_1,\, \mu_2) \le \rho_{\mathrm{K}}(\xi_1,\, \xi_2) \le \mathbb{E}\left(\|\xi_1 - \xi_2\|^{\,s}\right)^{1/(s+1)}\,. \tag{11}$$

Proof. Due to (10) we have

$$\mathbb{P}\left(\|\xi_1 - \xi_2\| \ge c\right) \le \frac{\mathbb{E}\left(\|\xi_1 - \xi_2\|^s\right)}{c^s}\,.$$

Solving the equation $c = \mathbb{E}\left(\|\xi_1 - \xi_2\|^s\right)/c^s$ concludes the proof. □

Remark 3.3. The bound in (11) leads to the following expectations:

- If ξ_1, ξ_2 are deterministic quantities, the right hand side of (11) converges to $\|\xi_1 - \xi_2\|$ for $s \to \infty$, and we obtain that $\rho_{\mathrm{P}}(\mu_1,\, \mu_2)$ and $\rho_{\mathrm{K}}(\xi_1,\, \xi_2)$ converge at least as fast as the expected value (As can be seen from (1) and (2) here even $\rho_{\mathrm{P}}(\mu_1,\, \mu_2) = \rho_{\mathrm{K}}(\xi_1,\, \xi_2) = \|\xi_1 - \xi_2\|$.).

- If the probability for large values of ξ_1, ξ_2 decays exponentially fast (as it is for instance the case for Gaussian random variables), all moments are finite, but they will grow with s. The relation from $\rho_{\mathrm{P}}(\mu_1,\, \mu_2)$ and $\rho_{\mathrm{K}}(\xi_1,\, \xi_2)$ to $\mathbb{E}\left(\|\xi_1 - \xi_2\|\right)$ becomes logarithmic (cf. [18, Ch. 6.2]).

- Finally, if not infinitely many moments are finite, but only moments up to some $s_0 < \infty$, the speeds in the two concepts can show significant differences (cf. the previous section and [18, Ch. 5]). This is for instance the case when ξ_1, ξ_2 are decaying polynomially, or come from a Lévy-distribution ([4, 24]).

In this work, we investigated properties of the Prokhorov and the Ky Fan metric, and pointed out connections and differences to other measures of convergence. For the Prokhorov metric, convergence (with rates) of Tikhonov regularized solutions of stochastic linear ill-posed problems was investigated in [13]. In [18] as well as in some forthcoming papers, these results of [13] have been extended to nonlinear problems, to more general regularization methods, and to the Ky Fan metric.

Finally, it should be mentioned that these concepts will also allow to obtain quantitative convergence results for estimators obtained by the Bayesian approach of inverse problems (see [20]).

Acknowledgments

The author thanks Prof. H.W. Engl and S. Kindermann for useful and stimulating discussions.

This work has been supported by the Austrian National Science Foundation FWF under project grant SFB F013/F1308.

References

[1] H. T. Banks and K. L. Bihari, *Modelling and estimating uncertainty in parameter estimation*, Inverse Problems, 17 (2001), pp. 95–111.

[2] R. G. Bartle, *The elements of integration*, John Wiley & Sons, Inc., New York, 1966.

[3] H. Bauer, *Probability theory and elements of measure theory*, Holt, Rinehart and Winston, Inc., 1972.

[4] J. Bertoin, *Lévy processes*, vol. 121 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 1996.

[5] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, Inc., 1968.

[6] N. Bissantz, T. Hohage, and A. Munk, *Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise*, Inverse Problems, 20 (2004), pp. 1773–1789.

[7] J. Davidson, *Stochastic limit theory*, The Clarendon Press Oxford University Press, New York, 1994.

[8] R. M. Dudley, *Distances of probability measures and random variables*, Ann. Math. Stat., 39 (1968), pp. 1563–1572.

[9] ——, *Real Analysis and Probability*, Wadsworth, second ed., 2003.

[10] B. Eisenberg and B. Ghosh, *A generalization of Markov's inequality.*, Stat. Probab. Lett., 53 (2001), pp. 59–65.

[11] J. Elstrodt, *Maß- und Integrationstheorie. (Measure and integration theory). 2., korr. Aufl.*, Berlin: Springer, 1999.

[12] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

[13] H. W. Engl, A. Hofinger, and S. Kindermann, *Convergence rates in the Prokhorov metric for assessing uncertainty in ill-posed problems*, Inverse Problems, 21 (2005), pp. 399–412.

[14] H. W. Engl and A. Wakolbinger, *On weak limits of probability distributions on Polish spaces.*, Stochastic Anal. Appl., 1 (1983), pp. 197–203.

[15] ——, *Continuity properties of the extension of a locally Lipschitz continuous map to the space of probability measures*, Monatsh. Math., 100, 2 (1995), pp. 85–103.

[16] K. Fan, *Entfernung zweier zufälliger Größen und die Konvergenz nach Wahrscheinlichkeit.*, Mathematische Zeitschrift, 49 (1944), pp. 681–683.

[17] A. L. Gibbs and F. E. Su, *On choosing and bounding probability metrics.*, Int. Stat. Rev., 70 (2002), pp. 419–435.

[18] A. Hofinger, *Ill-Posed Problems: Extending the Deterministic Theory to a Stochastic Setup*, vol. 51 of Schriften der Johannes-Kepler-Universität Linz. Reihe C: Technik und Naturwissenschaften, Trauner Verlag, Linz, 2006. (Thesis).

[19] P. Huber, *Robust Statistics*, John Wiley & Sons, Inc., 1981.

[20] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems.*, Applied Mathematical Sciences 160. New York, 2005.

[21] A. A. Markov, *Ischislenie Veroiatnostei (Calculus of Probabilities)*, Gosizdat, Moscow, 3rd ed., 1913.

[22] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, 1967.

[23] Y. V. Prokhorov, *Convergence of random processes and limit theorems in probability theory.*, Theor. Prob. Appl., 1 (1956), pp. 157–214.

[24] K.-I. Sato, *Lévy processes and infinitely divisible distributions*, vol. 68 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1999.

[25] G. Schay, *Nearest random variables with given distributions.*, Ann. Probab., 2 (1974), pp. 163–166.

[26] S. Smale and D.-X. Zhou, *Shannon sampling and function reconstruction from point values*, Bulletin of the American Mathematical Society, 41 (2004), pp. 279–305.

[27] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Stat., 36 (1965), pp. 423–439.

[28] D. Werner, *Funktionalanalysis. (Functional analysis).*, Berlin: Springer-Verlag, 1995.

[29] W. Whitt, *Preservation of rates of convergence under mappings.*, Z. Wahrscheinlichkeitstheor. Verw. Geb., 29 (1974), pp. 39–44.

[30] K. Yosida, *Functional analysis. 6th ed.*, Springer, Berlin, 1980.