Behavior and Convergence of Wasserstein Metric in the Framework of Stable Distributions*

VADYM OMELCHENKO

Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University Prague, Sokolovská 83, 186 75 Prague 8, Czech Republic.

Department of Econometrics, Institute of Information Theory and Automation, Pod Vodárenskou Věží 4, 182 08, Prague 8, Czech Republic.

e-mail: vadim224@yahoo.com

Abstract.
In this paper, we aim to explore the speed of convergence of the Wasserstein distance between stable cumulative distribution functions and their empirical counterparts. The theoretical results are compared with the results provided by simulations. The need to use simulations is explained by the fact that all the theoretical results which relate to the speed of convergence of the Wasserstein Metric in the set-up of stable distributions are asymptotic; therefore, the question of when that theory starts to be valid remains open. The asymptotic results are true only for relatively large numbers of observations exceeding hundreds of thousands. In cases dealing with lower numbers of observations, the speed of convergence turns out to be much slower than we might expect.

Keywords: Wasserstein Metric, Stable Distributions, Empirical Distribution Function.

JEL classification: C44
AMS classification: 90C15

---

1. Introduction

To define the Wasserstein Metric, let us introduce the following notation:

1. Let $\mathcal{P}(\mathbf{R}^s)$ denote the set of all Borel probability measures on $\mathbf{R}^s$,
2. $\mathcal{M}_1^p(\mathbf{R}^s) = \{\nu \in \mathcal{P}(\mathbf{R}^s) : \int_{\mathbf{R}^s} |z|_s^p \nu(dz) < \infty\}$,
3. Let $\mathcal{D}(P_F, P_G)$ be the set of such measures in , $\mathcal{P}(\mathbf{R}^s \times \mathbf{R}^s)$ whose marginal distributions are P_F and P_G,
4. $|\cdot|_s^2$ corresponds to the Euclidean norm; $|\cdot|_s^1$ - to the norm L_1 in $\mathbf{R}^s$.

*This research was supported by the Grant Agency of the Czech Republic under Grant 402/10/0956.

Definition 1. *Let F, G be two s-dimensional distribution functions corresponding to probability measures P_F and P_G, respectively. Then the value*

$$d_{W_s^p} := d_{W^p}(P_F, P_G) = \left(\inf_{\kappa \in \mathcal{D}(P_F, P_G)} \left\{ \int_{\mathbf{R}^s \times \mathbf{R}^s} |z - \hat{z}|_s^p \kappa(dz \times d\hat{z}) \right\} \right)^{1/p}$$

is called the Wasserstein Metric.

In the univariate case, for any continuous random variable we have the *Wasserstein Metric* equal to

$$d_{W_1^1}(P_F, P_G) = \int_R |F(z) - G(z)| dz,$$

(see [11])

Remark 2. According to [2] and [3], *Wasserstein Metric* determines the upper bound of the absolute differences between the optimal value of the real problem of stochastic programming and its empirical counterpart.

In the current paper, we will explore the rate of convergence of the Wasserstein Metric for stable distributions based on simulations, taking into account theoretical results in [1], [2] and [3].

2. Stable Distributions

Stable distributions are a broad family of statistical distributions allowing not only heavy tails but also skewness. They are called stable because their properties are preserved under convolution and defined as follows.

Definition 3. *A random vector $\mathbf{Y} = (Y_1, Y_2, ..., Y_s)$ is said to be a stable random vector in $\mathcal{R}^s$ if for any positive numbers a and b there is a positive number $c = c(a, b)$ and a vector $\mathbf{D} \in \mathcal{R}^s$ such that*

$$a \cdot \mathbf{Y}^{(1)} + b \cdot \mathbf{Y}^{(2)} =_d c \cdot \mathbf{Y} + \mathbf{D}$$

where $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are independent copies of $\mathbf{Y}$.

In the univariate case, stable distributions are characterized by four parameters, one of which determines how heavy tailed the distribution is; it is called the tail index and will be denoted by α. The tail index lies in the interval $(0, 2]$ and lower values of α correspond to heavier tails. Normal distribution is a special case of a stable distributions; it corresponds to $\alpha = 2$. If a stable distribution Y has a tail index $\alpha < 2$, then for any $a \geq \alpha$ we have $E|Y|^a = \infty$, if $a \in (0, \alpha)$ then $E|Y|^a < \infty$. The parameter σ is called scale parameter, and $\sigma > 0$. The parameter μ is the shift parameter and $\mu \in \mathbf{R}$. The last parameter is skewness β and $\beta \in [-1, 1]$. Stable distributions have widespread applications in economics and finance (see [5]). They are mostly used to describe the behaviour of highly volatile data having a much higher kurtosis than that allowed by normal models. The assumption that Y is normal is usually justified by the Central Limit Theorem. The analogous assumption that Y is stable is justified by the Generalized Central Limit Theorem, which differs from the Central Limit Theorem by dropping the assumption of finiteness of the mean and variance. If we want to include heavy tails in the model without completely ruling out normality, the best option is to use stable distributions due to their form of characteristic function, as well as their limit and convolution properties. The defining characteristics, and the reason for the term *stable*, is that they retain their shape under convolution (see [5]): if $Y, Y_1, Y_2, ..., Y_n$ are i.i.d. stable random variables, then for every $n \in \mathbf{N}$

$$Y_1 + Y_2 + ... + Y_n =_d c_n Y + d_n$$

for certain constants $c_n > 0$ and d_n (see [6]). The Generalized Central Limit Theorem states that the only possible distributions with a domain of attraction are stable (see [6]). Similar reasons for utilizing the normal distribution hold for the stable distributions that is why they have a wide range of applications.

Remark 4. The characteristic function of a stable distribution is given explicitly for the univariate case, but neither the densities nor distribution functions of stable distributions can be expressed in terms of elementary functions, except for Lévy, Cauchy and Normal distributions (see [4]).

2.1. Some definitions and basic theorems

Before turning our attention to the main topic, we need to define and recall some related basic definitions, notations, and theorems.

Definition 5. *Functions $f, g : \mathbf{R} \to \mathbf{R}$ are said to be asymptotically equal if* $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Remark 6. We denote asymptotic equality with the symbol $\approx$.

Theorem 7. *[5] For a stable distribution $Y \sim S_\alpha(\sigma, 0, 0)$ with $\alpha < 2$ the density function f and distribution function F have the form*

$$f(y) \approx \frac{\sigma^\alpha \sin(\pi\alpha/2)\Gamma(\alpha+1)/\pi}{|y|^{\alpha+1}} \quad \text{as} \quad |y| \to \infty,$$

$$1 - F(y) \approx \frac{(\alpha+1)\sigma^\alpha \sin(\pi\alpha/2)\Gamma(\alpha+1)/\pi}{y^\alpha} \quad \text{as} \quad y \to \infty,$$

$$F(-y) \approx \frac{(\alpha+1)\sigma^\alpha \sin(\pi\alpha/2)\Gamma(\alpha+1)/\pi}{|y|^\alpha} \quad \text{as} \quad y \to \infty.$$

Proof. see [5]. □

Theorem 8. *[4] All stable distributions functions are absolutely continuous and have densities.*

Proof. (see [4],pp 21-24). □

Simulation of a symmetric stable random variable can be conducted as follows:

Theorem 9. *[9] Suppose that γ has uniform distribution on $(-\pi/2, \pi/2)$, and W is exponentially distributed with unit mean. Let us also suppose that γ and W are independent, then*

$$Y = \frac{\sin \alpha\gamma}{(\cos \gamma)^{1/\alpha}} \left(\frac{\cos((1-\alpha)\gamma)}{W} \right)^{(1-\alpha)/\alpha}$$

has a stable distribution given by $Y \sim S_\alpha(1, 0, 0)$

Proof. See [9]. □

Remark 10. Theorem 8 enables us to simulate $Y_0 \sim S_\alpha(1, 0, 0)$ and it may be easily shown that $\sigma \cdot Y_0 + \mu \sim S_\alpha(\sigma, 0, \mu)$.

3. Features of Wasserstein Metric

In this section, we mention the basic properties of the Wasserstein Metric in the framework of stable distributions, and heavy tails in general. By the notion *heavy tailed distribution*, we will understand the distribution of a random variable X for which there exists $p > 0$ such that $E|X|^p = \infty$. Let us introduce a system of assumptions:

A1. $\{\xi^i\}_{i=1}^{\infty}$ is an independent random sequence corresponding to F,

A2. The marginal distributions P_{F_i}, $i = 1, 2, ..., s$ are absolutely continuous w.r.t the Lebesgue measure on $\mathbf{R}^1$.

Definition 11. *The sequence $\zeta_n, n \in \mathbf{N}$ is said to be stochastically bounded if for an arbitrary $\epsilon > 0$ there exists a number $c > 0$ such that for all $n \in \mathbf{N}$*

$$P(|\zeta_n| > c) < \epsilon.$$

Theorem 12. *[1] Let $s = 1$, $\{\xi^i\}_{i=1}^{\infty}$, $N = 1, 2, ...$ be a sequence of independent random values corresponding to a heavy tailed distribution F with the shape parameter $\alpha \in (1, 2)$. Then the sequence*

$$\frac{N}{N^{1/\alpha}} \int_{\infty}^{\infty} |F^N(z) - F(z)| dz, \quad N = 1, 2, ...$$

is stochastically bounded if and only if

$$\sup_{t>0} t^\alpha P\{\omega : |\xi| > t\} < \infty.$$

Proof. See [1]. □

Theorem 13. *[3] Let assumptions A1. and A2. be fulfilled, $s = 1$, $P_F \in \mathcal{M}_1^1(\mathbf{R}^1)$. If*

$$\int_{-\infty}^{\infty} \sqrt{F(z)(1 - F(z))} < +\infty$$

then

$$\sqrt{N} \int_{-\infty}^{\infty} |F^N(z) - F(z)| dz \longrightarrow_d \int_{-\infty}^{\infty} |\mathbf{U}(F(z))| dz,$$

where $\mathbf{U}$ *denotes the Brownian bridge.*

Proof. See [3]. □

Theorem 14. *[1]*

1. *If $s = 1$, F is a distribution function, f a density function of a stable distribution with $1 < \alpha < 2$, and F^N – an empirical distribution function determined by N i.i.d. random variables $\xi_1, \xi_2, ..., \xi_N$ having the distribution function F, then*

$$\frac{N}{N^{1/\alpha}} \int_{-\infty}^{\infty} |F^N(z) - F(z)| dz \longrightarrow_d \frac{1}{\alpha}\left(\frac{c_1}{\alpha}\right)^{1/\alpha} \int_0^{\infty} |N_1(s) - s| s^{-1-1/\alpha} ds +$$

$$+ \frac{1}{\alpha}\left(\frac{c_2}{\alpha}\right)^{1/\alpha} \int_0^{\infty} |N_2(s) - s| s^{-1-1/\alpha} ds$$

where $N_i, i = 1, 2$ are independent Poisson processes with an intensity of 1,

$$c_1 = \lim_{x \to \infty} f(x) x^{\alpha+1} \quad , and \quad c_2 = \lim_{x \to -\infty} f(x) |x|^{\alpha+1}.$$

2. $$\mathbf{E}\left(\frac{N}{N^{1/\alpha}} \int_{-\infty}^{\infty} |F^N(z) - F(z)| dz\right) \longrightarrow \mathbf{E} \int_{-\infty}^{\infty} |K(z) - \mathbf{E}K(z)| dz$$

where

$$K(z) - \mathbf{E}K(z) = N_1\left(\frac{c_1}{\alpha t^\alpha}\right) - \frac{c_1}{\alpha t^\alpha}, \quad t > 0$$

$$K(z) - \mathbf{E}K(z) = N_2\left(\frac{c_2}{\alpha t^\alpha}\right) - \frac{c_2}{\alpha t^\alpha}, \quad t < 0$$

and $K(0) - \mathbf{E}K(0) = 0$.

Proof. See [1] pp 1011-1027. □

Notation. WM denotes the Wasserstein Metric. Sometimes, it will be necessary to emphasize the dependence of *Wasserstein Metric* not only on N but also on α therefore, in these cases we will denote it by WM(N, α). We will also deal with the value $N^{p(1-1/\alpha)} \int_{\mathbf{R}} |F_N(z) - F(z)| dz$ where $p \in (0, 1]$ and we shall denote this value for $p \in [0, 1]$ by SWM(p) where SWM means *Scaled Wasserstein Metric.* In cases when it is necessary to emphasize the dependence not only on p but also on N, it will be denoted by SWM(p, N).

Theorem 15 (Limit Comparison Test).
Let us suppose that:

1. $f, g : [A, \infty) \to \mathbf{R}^+$,
2. f *and* g *are continuous on* $[A, \infty)$,
3. $\lim_{z\to\infty} \frac{f(z)}{g(z)} = L > 0$,

where L *and* A *are finite positive numbers. Then both* $\int_A^\infty f(z)dz$ *and* $\int_A^\infty g(z)dz$ *either converge or diverge.*

Proof. See [7]. □

Theorem 16. *[10] Let* ζ_n, $n \in \mathbf{N}$ *and* ζ *be random variables in* $\mathbf{R}$, *all of which defined in the probability space* (Ω, A, P). *Then* $p\text{-}\lim_{n\to\infty} \zeta_n = \zeta$ *implies* $d\text{-}\lim_{n\to\infty} \zeta_n = \zeta$ *where* $p\text{-}\lim$ *and* $d\text{-}\lim$ *means convergence in probability and distribution, respectively.*

Proof. See [8]. □

Theorem 17. *Let* s *be 1 and assumptions A1. and A2. be fulfilled,* $\gamma > 0$ *a constant,* F *a distribution function of a stable distribution with* $\alpha < 2$, F^N *an empirical distribution function derived from* F, $P_F \in \mathcal{M}_1^1(\mathbf{R}^1)$ *and* $s = 1$. *If one of the equivalent conditions*

$$\int_{-\infty}^{\infty} \sqrt[\gamma]{F(z)(1-F(z))}dz < +\infty \quad \text{or} \quad \gamma < \alpha$$

holds true then

$$\frac{N}{N^{1/\gamma}} \int_{-\infty}^{\infty} |F^N(z) - F(z)|dz \longrightarrow_d 0$$

Proof.

1. Let us prove first that $\int_{-\infty}^{\infty} \sqrt[\gamma]{F(z)(1-F(z))}dz < +\infty$ implies that $\gamma < \alpha$. By Theorem 7, we get $F(z) \approx 1 - \frac{c}{z^\alpha}$ for a certain constant $c > 0$. Hence by The Limit Comparison Test, we have for any positive number A that $\int_A^\infty (F(z)(1-F(z)))^{1/\gamma}dz$ converges whenever the integral $\int_A^\infty \left(\frac{c}{z^\alpha}\right)^{1/\gamma} \left(1-\frac{c}{z^\alpha}\right)^{1/\gamma} dz$ converges. We have

$$\int_A^\infty \left(\frac{c}{z^\alpha}\right)^{1/\gamma} \left(1-\frac{c}{z^\alpha}\right)^{1/\gamma} dz = c^{1/\gamma} \int_A^\infty \left(\frac{z^\alpha - c}{z^{2\alpha}}\right)^{1/\gamma} dz$$

Applying The Limit Comparison Test we get that the integral $c^{1/\gamma} \int_A^\infty \left(\frac{z^\alpha - c}{z^{2\alpha}}\right)^{1/\gamma} dz$ converges whenever the integral $c^{1/\gamma} \int_A^\infty \left(\frac{z^\alpha}{z^{2\alpha}}\right)^{1/\gamma} dz$ is convergent. We have

$$c^{1/\gamma} \int_A^\infty \left(\frac{z^\alpha}{z^{2\alpha}}\right)^{1/\gamma} dz = c^{1/\gamma} \int_A^\infty \frac{1}{z^{\alpha/\gamma}} dz$$

Analogously, we will have that $\int_{-\infty}^{-A}(F(z)(1-F(z)))^{\gamma}dz$ converges whenever the integral $c^{1/\gamma}\int_{-\infty}^{-A}\frac{1}{|z|^{\alpha/\gamma}}dz$ converges. Both of these integrals will converge only for $\alpha/\gamma > 1$ i.e. $\gamma < \alpha$. The integral $\int_{-A}^{A}(F(z)(1-F(z)))^{1/\gamma}dz$ obviously converges because $F(z)$ is bounded and absolutely continuous on $\mathbf{R}$.
Hence, the integral $\int_{-\infty}^{\infty}(F(z)(1-F(z)))^{1/\gamma}dz$ will converge because each of the three integrals converges. The opposite implication also follows from the Limit Comparison Test.

2. Due to $(1-F(z)) \approx c\frac{1}{z^{\alpha}}$ with $c > 0$, it is clear that $\lim_{z\to\infty} z^{\alpha}P\{|\xi| > z\} < \infty$. Therefore $N/N^{1/\alpha}\int_{\mathbf{R}}|F_n(z)-F(z)|dz$ is stochastically bounded by Theorem 12. Due to the fact that $\gamma < \alpha$ and the stochastic boundedness of

$$\frac{N}{N^{1/\alpha}}\int_{\mathbf{R}}|F_N(z)-F(z)|dz,$$

we have

$$\frac{N}{N^{1/\gamma}}\int_{\mathbf{R}}|F_N(z)-F(z)|dz = \frac{N^{1/\alpha}}{N^{1/\alpha}}\frac{N}{N^{1/\gamma}}\int_{\mathbf{R}}|F_N(z)-F(z)|dz =$$
$$= \frac{1}{N^{1/\gamma-1/\alpha}}\left(\frac{N}{N^{1/\alpha}}\int_{\mathbf{R}}|F_N(z)-F(z)|dz\right).$$

Due to Markov inequality, we have for any $\epsilon > 0$ and $N \in \mathbf{N}$

$$P\left(\frac{\left(\frac{N}{N^{1/\alpha}}\int_{\mathbf{R}}|F_N(z)-F(z)|dz\right)}{N^{1/\gamma-1/\alpha}} > \epsilon\right) \le \frac{1}{N^{\frac{1}{\gamma}-\frac{1}{\alpha}}}\frac{\mathbf{E}\left(\frac{N}{N^{1/\alpha}}\int_{\mathbf{R}}|F_N(z)-F(z)|dz\right)}{\epsilon}$$

Taking Theorem 14, we can see that the expected value in the right hand side of the inequality is always finite and $1/N^{1/\gamma-1/\alpha}$ converges to zero as N tends to infinity. This proves convergence in probability and by Theorem 16, convergence in probability implies convergence in distribution which proves the statement.

□

Corollary 18. *Let us take the assumptions of Theorem 17. Then*

$$d\text{-}\lim_{N\to\infty}\mathrm{SWM}(p,N) = 0$$

whenever $p < 1$.

Proof. We have

$$\frac{N}{N^{1/\gamma}}\int_{\mathbf{R}}|F_n(z)-F(z)|dz = N^{1-1/\gamma}\int_{\mathbf{R}}|F_n(z)-F(z)|dz =$$
$$= N^{(1-1/\alpha)\frac{1-1/\gamma}{1-1/\alpha}}\int_{\mathbf{R}}|F_n(z)-F(z)|dz = \mathrm{SWM}\left(\frac{1-1/\gamma}{1-1/\alpha}, N\right).$$

For $\gamma > 0$ and $\alpha \in (0,2]$ the ratio $\frac{1-1/\gamma}{1-1/\alpha}$ is smaller than 1 whenever $\gamma < \alpha$, and hence the statement follows from Theorem 17. □

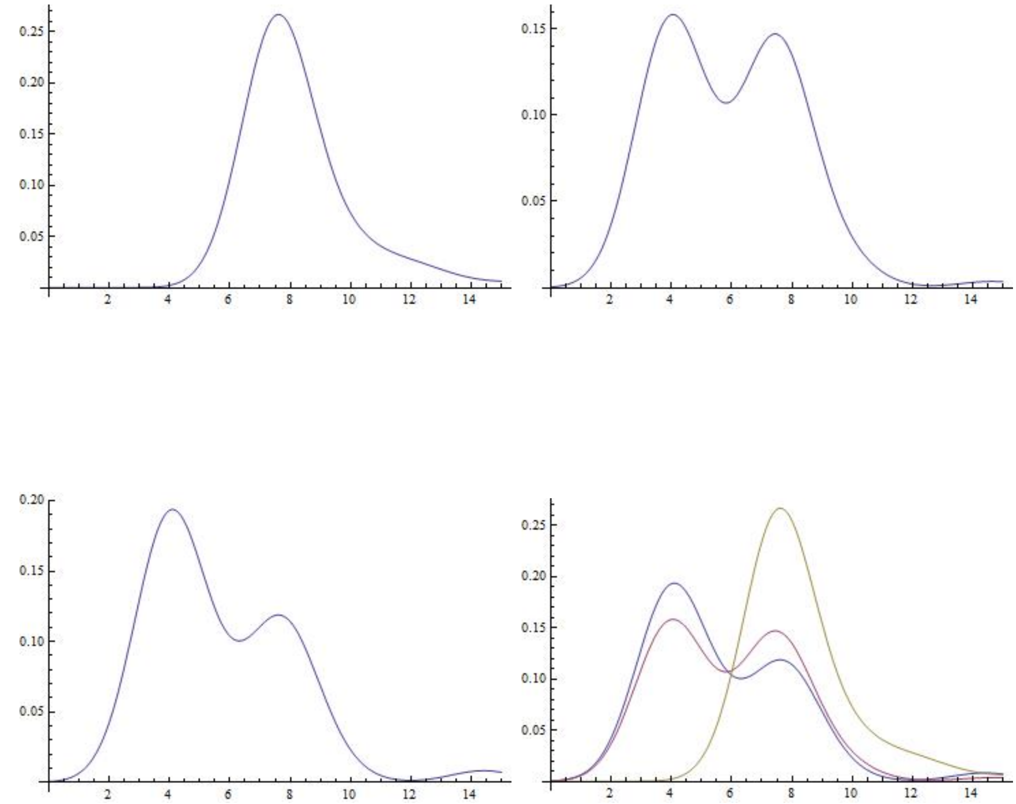

Figure 1: Density functions of SWM(1) for $\alpha = 1.1$ where N=200 (top left), N=40000 (top right), N=80000 (bottom left and all the graphs combined - bottom right)

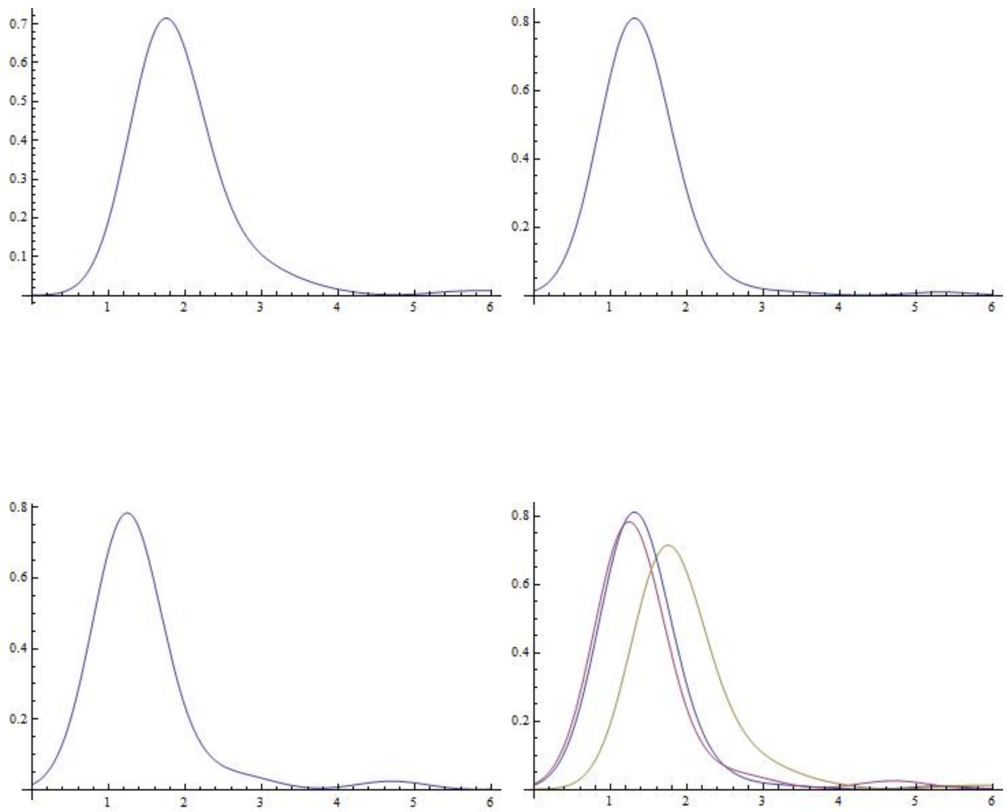

Figure 2: Density functions of SWM(0.75) for $\alpha = 1.5$ where N=200 (top left), N=80000 (top right), N=160000 (bottom left) and all the graphs combines - bottom right)

4. Simulation

The rate of convergence is explored by simulating random samples from $S_\alpha(1,0,0)$ to construct the empirical distribution function and calculate the Wasserstein Metric. In the following two subsections, we will perform an analysis of the density function of the distribution of SWM(1) and apply a regression analysis to the Wasserstein Metric to study its dependence on N. The reason we explore only the special case when $\sigma = 1$, $\mu = 0$ and $\beta = 0$ is the fact that for any σ and μ a proper shifting and scaling will transform the initial stable distribution into the one with σ equal to 1 and μ equal to 0, i.e. if $Y \sim S_\alpha(\sigma, \mu, \beta)$ then $\frac{Y-\mu}{\sigma} \sim S_\alpha(1, 0, \beta)$. It is clear that shifting does not affect the form of the distributions tail. The scaling also does not affect the rate of convergence because if we have the distribution function F for $Y_0 \sim S_\alpha(1,0,0)$ then appealing to the basic properties of stable distributions, we get for $Y \sim S_\alpha(\sigma, 0, 0)$ with $\sigma > 0$ that $Y =_d \sigma Y_0$ and the distribution function of Y will be $F_Y(y) = F\left(y/\sigma\right), y \in \mathbf{R}$.

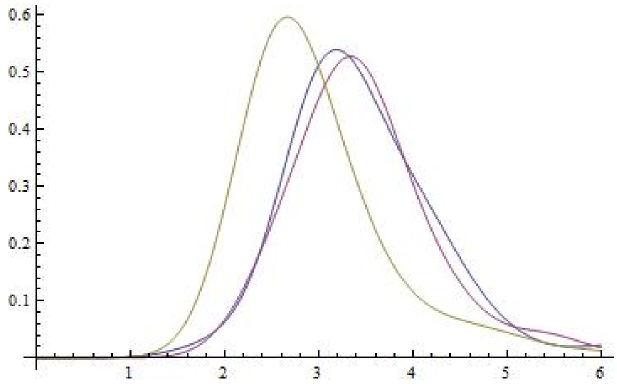

Figure 3: Density functions of SWM(1) for $\alpha = 1.5$ where N=200, 80000 and 160000

4.1. Density Functions and Sample Characteristics of SWM(1)

To estimate the density function of the *Scaled Wasserstein Metric*, we used kernel estimates of the density with the kernel $K(\cdot)$ equal to the density of the normal distribution, i.e. $K(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$, $z \in \mathbf{R}$. For $\alpha = 1.1, 1.2, ..., 1.9, 2$, we calculated the sample medians, mean and density functions for samples sizes of SWM(1) equal to 20 where N ranged to 200, 40000 and 80000 and 160000. For $\alpha = 1.1$, the graph of the kernel estimate of the density function for SWM(1) can be seen in Figure 1. To test equality in the distribution of SWM(1) for different number of observations N, we will apply the Kolmogorov-Smirnov test [8]. This procedure is carried out because SWM(1) converges in distribution. It is clear from Figure 1 that the density functions differ substantially and Kolmogorov Smirnov test rejects the null hypothesis of equality in distribution for the samples of SWM(1) with $N = 40000$ and $N = 80000$ on 95% confidence level. But for $\alpha = 1.1$ the higher the number of observations, the closer to zero the shape of the density function. Moreover, the median of SWM(1) also decreases for larger numbers of observations being 7.612 for $N = 200$, 5.551 - for $N = 40000$, 5.478 - for $N = 80000$ and 5.481 - for $N = 160000$. For any $\alpha > 1.2$ the analogous medians of SWM(1) form a slightly rising sequence so that the higher N, the higher the median.
For each $\alpha = 1.1, 1.2, ..., 1.9, 2$, the Kolmogorov Smirnov Statistic was calculated for 2 samples with 20 elements of SWM(1) with $N = 160000$ and SWM(1) with $N = 80000$ reaching its highest value 0.46 at $\alpha = 1.5$. Because in each case the value of the Kolmogorov-Smirnov statistic falls below 0.46, we can accept on 0.95% confidence level the null hypothesis of equality in distribution for these two samples. In Figure 2, we can observe the kernel estimate of the density of SWM(0.75) for

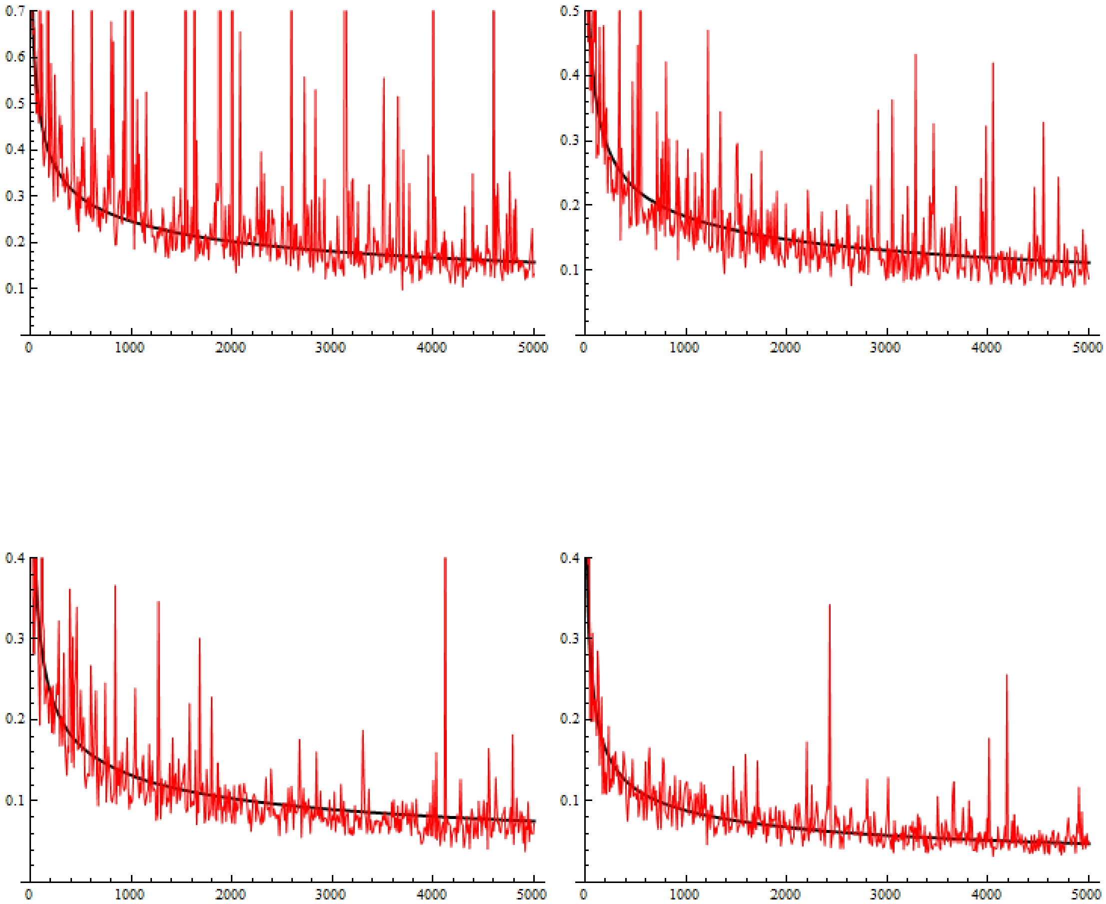

Figure 4: Convergence of Wasserstein Metric for $\alpha = 1.5$ (top left), $\alpha = 1.6$ (top right) $\alpha = 1.7$ (bottom left), $\alpha = 1.8$ (bottom right) where N ranges from 10 to 5000 with step 10.

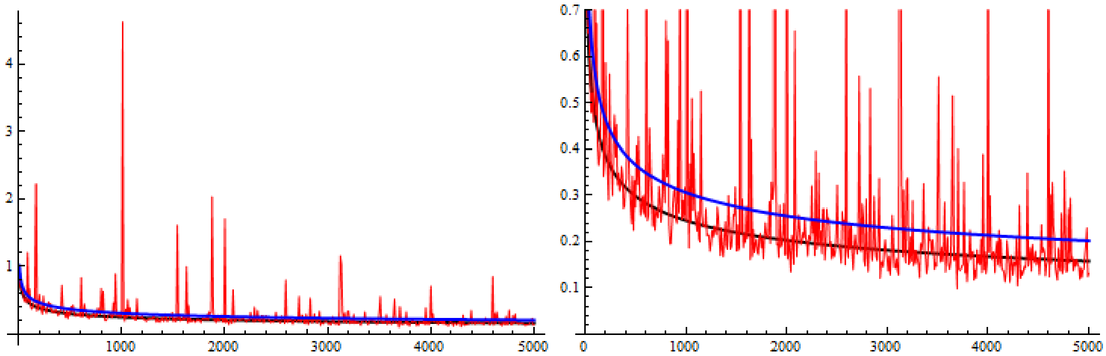

Figure 5: Comparison of the fit by classical and median regression for $\alpha = 1.5$. The whole picture (left) and the same zoomed picture (right).

Table 1: Convergence rate following from regression analysis

α	Percentage of $(1 - 1/\alpha)$ lower bound	Rate of convergence up to 5000	Rate of convergence of medians
$\alpha = 1.10$	134%	0.1218	0.1100
$\alpha = 1.20$	107%	0.1783	0.1819
$\alpha = 1.30$	90%	0.2076	0.1992
$\alpha = 1.40$	83%	0.2371	0.2412
$\alpha = 1.50$	75%	0.2500	0.2679
$\alpha = 1.60$	78%	0.2925	0.3018
$\alpha = 1.70$	81%	0.3335	0.3212
$\alpha = 1.80$	86%	0.3822	0.4013
$\alpha = 1.90$	90%	0.4263	0.4215
$\alpha = 1.95$	95%	0.4628	0.4713
$\alpha = 2.00$	100%	0.4992	0.4912

$\alpha = 1.5$ and see that the shape of the empirical density functions accumulates closer to zero for higher numbers of observations. The analogous picture can be observed for any other $\alpha = 1.1, 1.2, .., 1.9, 2$. In Figure 3 we can see that the convergence of SWM(1) for $\alpha = 1.5$ obtained from the same data points as SWM(0.75) is much different: in this case the higher N, the further the shape of the empirical density is from zero. But when $\alpha = 2$, i.e., in the case of a normal distribution, we can observe the stagnation of the SWM(1) for much lower values of N.

4.2. Regression Analysis

To explore the speed of convergence of the Wasserstein Metric, we will fit the series of the Wasserstein Metric values to the model

$$WM(N, \alpha) \sim \frac{m}{N^b}$$

because according to Theorem 17 and Corollary 18, Wasserstein Metric follows the model with $b = (1 - 1/\alpha)$, i.e.

$$WM(N, \alpha) = O_P\left(\frac{1}{N^b}\right) \quad \text{with} \quad b = 1 - 1/\alpha.$$

We will choose the rate $b = b(\alpha)$ and coefficient $m = m(\alpha)$ for which the median of the sample $|WM(k, \alpha) - \frac{m}{k^b}|, |WM(k+1, \alpha) - \frac{m}{(k+1)^b}|, ..., |WM(N, \alpha) - \frac{m}{N^b}|$ is lowest for $k, N \in \mathbf{N}$, $N > k$. We estimate the rate of convergence using a series with fixed α and varying N and we choose the value of k which minimizes the overall median but its value will be restricted by $10\%N$ so that the number of excluded observations will not be higher than 10%. It yields the following problem:

$$\min_{m>0, b>0, k\in\mathbf{N}, k\leq 0.1\cdot N} \sum_{i=k}^{N} \left|WM(i, \alpha) - \frac{m}{i^b}\right|$$

with unknown b and m. It is actually L_1 regression. The rates will be determined for 10 values of α and the rest can be obtained by linear interpolation. In our settings, N varies from 10 to 5000 with the step equal to 10, i.e. $N \in \{10n : n = 1, 2, 3, ..., 500\}$. In Figure 4, we can observe the Wasserstein Metric values for $\alpha = 1.5$, $\alpha = 1.6$, $\alpha = 1.7$ and $\alpha = 1.7$. For $\alpha = 1.5$ and $\alpha = 1.6$, the sequence of Wasserstein Metric values was fitted to functions $f_{1.5}(t) = \frac{1.66982}{t^{0.25}}$ and $f_{1.6}(t) = \frac{1.7983}{t^{0.2825}}$, respectively. According to the calculations of regression analysis, a higher rate of convergence can be obtained than the upper bound given by $1 - 1/\alpha$ only for $\alpha = 1.1$ or 1.2 $N < 100000$. For $\alpha = 1.5$, the convergence rate is about $0.75 \cdot (1 - 1/\alpha)$. For $\alpha = 1.9$, it is about $0.95 \cdot (1 - 1/\alpha)$ and it is again $(1 - 1/\alpha)$ for $\alpha = 2$. Table 1. summarizes the results of the regression analysis. When $\alpha > 1.2$, the speed of convergence is lower than it should be according to the theory. When $\alpha > 1.7$, the results provided by classical linear regression are almost the same and yield r-squared larger than 85%. Otherwise the difference will be substantial. In Figure 5, the difference in fitting the Wasserstein Metric by classical and median regression can be observed. But the estimate of the speed by the classical regression is 0.268 and by the median regression - 0.25. Analogous regression analysis was applied on medians of the Wasserstein Metric values for α ranging from 1.1 to 2 with a step of 0.1 and the number of observations taking the following values: 200, 400, 600, 800, 1200, 1400, 1800, 2200, 2600, 3000, 5000, 40000. For each pair of α and N the median was calculated from 20 simulated Wasserstein Metric values and the same regression analysis was applied on those values. The results of this analysis are summarized in the fourth column of Table 1.

Remark 19. In addition to calculating the Wasserstein Metric for stable distributions, we were making analogous calculations for Pareto distributions and we obtained that the rate of their convergence was in accordance with Theorem 10 without substantial deviations from it for relatively low values of N. Table 2 presents the calculations of the convergence rates of of the Wasserstein Metric values of Pareto distributions with the distribution function $F(z) = 1-1/z^{\alpha}$ in the same set-up as in the case of stable distributions referring to Table 1, i.e., $N \in \{10 \cdot n : n = 1, 2, 3, .., 500\}$ where $\hat{b}(\alpha)$ is the estimate of the rate of convergence from the data.

Table 2: Convergence rate of Pareto distribution

α	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
$(1-1/\alpha)$	0.09	0.16	0.23	0.28	0.33	0.37	0.41	0.44	0.47
$\hat{b}(\alpha)$	0.011	0.143	0.258	0.275	0.326	0.363	0.402	0.461	0.474

Conclusions

The present paper deals with the investigation of convergence rates of the Wasserstein Metric values in the framework of stable distributions by running simulations and checking convergence rates based on two analyses: that of analysis of density functions and regression analysis. In spite of the fact that the stable distributions have Paretian tails, the convergence of their Wasserstein Metric values differs from the convergence of the Pareto distribution. For the numbers of observations N below 100000 and $\alpha > 1.2$, the speed of convergence for the Wasserestein metric is lower than the one given by the theoretical results; in the case of $\alpha = 1.5$, it is only 75% of the theoretical rate. The speed of convergence given by the theoretical results is attained for the number of observations N larger than 100000.

Acknowledgements

The author wishes to thank Vlasta Kaňková for insightful reading the manuscript and many helpful comments.

References

[1] Barrio, E., Giné, E., and Matrán, C.: Central limit theorems for the Wasserstein distance between the empirical and true distributions. *Annals of Probability* 27 (1999), 2, 1009-1071.

[2] Kaňková, V.: Empirical Estimates in Economic and Financial Problems via Heavy Tails. In: *Proceedings of the 30-th International Conference Mathematical Methods in Economics 2012* (J. Remík M. Hušková, and D. Stavárek z, eds.), Silesian University in Opava, School of Business Administration in Karviná (The Czech Republic), 2012, 396–401.

[3] Kaňková, V., and Houda, M.: Empirical Estimates in Stochastic Programming. In: *Proceedings of the Prague Stochastics 2006* (M. Hušková, and M. Janžura, eds.), MATFYZPRESS, Prague, 2006, 426–436.

[4] Klebanov, L.: *Heavy tailed distributions.* Matfyzpress, Prague, 2003.

[5] Nolan, J. P.: Modeling Financial Data With Stable Distributions. In: *Handbook of Heavy Tailed Distributions in Finance* (S. Rachev, eds.), Elsevier Science B.V. 2003

[6] Omelchenko, V.: Stable distributions and applications in finance. Diploma Thesis (supervisor L. Klebanov) Faculty of Mathematics and Physics, Charles University, Prague 2007.

[7] Piskunov N.: *Differential and Integral Calculus.* Mir Publishers, Moscow, 1989.

[8] Réniy A.: *Probability Theory (in Czech).* , Academia, Prague 1972.

[9] Samorodnitsky, G. and Taqqu, M.: *Stable Non-Gaussian Random Processes*, Chapman and Hall, New York. (1994).

[10] Stepan, J. : *Theory of Probability - Mathematical Foundation*, Academia, Praha. (1994).

[11] Vallander, S. S.: Calculation of the Wasserstein distance between probability distributions on the line, *Theory of Probability and its Applications* **18** (1973), 784–786.