Russian Math. Surveys **71**:6 1021–1079 *Uspekhi Mat. Nauk* **71**:6 37–98

DOI: https://doi.org/10.1070/RM9749

Proximity of probability distributions in terms of Fourier–Stieltjes transforms

S. G. Bobkov

Abstract. A survey is given of some results on smoothing inequalities for various probability metrics (in particular, for the Kolmogorov distance), and some analogues of these results in the class of functions of bounded variation are presented.

Bibliography: 61 titles.

Keywords: probability metrics, smoothing inequalities.

Contents

This work was carried out with the support of the Alexander von Humboldt Foundation and the National Science Foundation (grant NSF DMS-1612961).

AMS 2010 *Mathematics Subject Classification*. Primary 60E05, 60E10; Secondary 60B10, 60F05.

1. Introduction

Let $\mathscr{F}$ denote the space of all distribution functions on the real line, that is, non-decreasing right-continuous functions $F\colon \mathbb{R} \to [0,1]$ such that $F(-\infty) = 0$ and $F(\infty) = 1$. There is a variety of popular metrics and pseudometrics d on $\mathscr{F}$, and questions of estimating the distance $d(F,G)$ arise occasionally in one or another approximation problem (for example, in the theory of summation of independent random variables). Smoothing inequalities allow one to estimate $d(F,G)$ in terms of the corresponding characteristic functions (Fourier–Stieltjes transforms)

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x), \quad g(t) = \int_{-\infty}^{\infty} e^{itx}\, dG(x), \qquad t \in \mathbb{R}, \tag{1.1}$$

for which additional smoothness-type conditions on G are possibly required.

A very important example is the uniform distance (the Kolmogorov distance)

$$\rho(F,G) = \sup_x |F(x) - G(x)|.$$

Integrating (1.1) by parts, we obtain the equality

$$\frac{f(t) - g(t)}{-it} = \int_{-\infty}^{\infty} e^{itx} (F(x) - G(x))\, dx, \qquad t \neq 0,$$

where the integral must in general be understood as the principal value. If the function $(f(t) - g(t))/t$ is integrable, then the function $F - G$ must be uniformly continuous, and we can apply the inverse Fourier transform:

$$F(x) - G(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{f(t) - g(t)}{-it}\, dt. \tag{1.2}$$

This gives us the elementary estimate

$$\rho(F,G) \leqslant \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{f(t) - g(t)}{t} \right| dt, \tag{1.3}$$

which is valid without any smoothness assumptions.

However, in many problems the integral on the right-hand side diverges, so that this estimate turns out to be useless. Nevertheless, under certain smoothness conditions and at the expense of a small error, the integral in (1.3) can be replaced by the integral over a finite interval $[-T,T]$, on which we know or can show that f and g are sufficiently close. This is achieved by smoothing the distributions F

and G with the help of a suitable (specially chosen) distribution H that depends on the parameter T. Namely, we consider convolutions

$$(F * H)(x) = \int_{-\infty}^{\infty} F(x-y)\, dH(y)$$

and obtain relations (smoothing inequalities) of the form

$$\rho(F,G) \leqslant c\rho(F*H, G*H) + \varepsilon,$$

where c is usually an absolute constant and ε can depend on T and some other parameters determined by F and/or G.

Similar relations are also studied for other metrics d. Finally, estimates for $d(F,G)$ are derived in terms of the proximity of f to g; such estimates are also called smoothing inequalities (see, for example, [1]–[3]). In this survey we look at some results on this topic, and we discuss standard approaches to them and possible refinements or generalizations, but we certainly do not claim completeness of our exposition. We also present several new inequalities for distances such as the quadratic Kantorovich distance, the distance in the L^1-metric, and the Kolmogorov distance (under additional smoothness conditions and also with a polynomial weight). A number of results on the rate of convergence in the central limit theorem for weak metrics (that is, metrics responsible for convergence of probability distributions in the weak topology) is presented at the end of the survey.

2. Kolmogorov distance

The first result on proximity of distributions in terms of the corresponding Fourier–Stieltjes transforms was obtained by Esseen in his classical work [4], where the following important theorem was proved.

Theorem 2.1. *If a distribution function G is differentiable and $|G'(x)| \leqslant L$ for all $x \in \mathbb{R}$, then for any $T>0$ and all $b>1/(2\pi)$*

$$\rho(F,G) \leqslant b \int_{-T}^{T} \left| \frac{f(t)-g(t)}{t} \right| dt + c(b)\frac{L}{T}\,, \tag{2.1}$$

where the constant $c(b)$ depends only on b.

The main ideas for proving this theorem appeared in the earlier paper [5] by Berry, who considered the particular case of the standard normal distribution function $G=\Phi$. Hence the inequality (2.1) is called the Berry–Esseen inequality, as is the estimate of the convergence rate in the central limit theorem that is derived from it (Theorem 20.1). The inequality (2.1) also remains valid for a broader class of functions: F can be an arbitrary bounded non-decreasing function, while G can be a function of bounded variation, with $F(-\infty)=G(-\infty)=0$ and $|G'(x)| \leqslant L$ (the monotonicity condition for G is dropped [2]–[4]). The definition of the uniform distance remains the same:

$$\rho(F,G) = \sup_x |F(x)-G(x)|.$$

Such a generalization allows us to work with Edgeworth expansions for distributions of sums of independent random variables.

The estimate (2.1) (or a similar relation) can be derived on the basis of the smoothing inequality

$$\rho(F,G) \leqslant \frac{1}{1-2\gamma}\rho(F*H_T, G*H_T) + 2l\frac{1-\gamma}{1-2\gamma}\,\frac{L}{T}\,, \tag{2.2}$$

where $H_T(x) = H(Tx)$ and H is the probability distribution with density

$$\psi(x) = \frac{1}{2\pi}\left(\frac{\sin(x/2)}{x/2}\right)^2 = \frac{1-\cos x}{\pi x^2}\,, \tag{2.3}$$

where the parameter $l > 0$ can be any number satisfying the condition $1-\gamma = H[-l,l] \equiv H(l) - H(-l) < 1/2$ (see the proof of Lemma 16.1).

The distribution function H with the density (2.3) and convolution powers of it are most popular in smoothing inequalities. It has the triangular characteristic function $h(t) = (1-|t|)_+$, so the characteristic function of the distribution H_T has the compact support $[-T,T]$. Applying (1.3) to the smoothed distributions, we arrive at the estimate

$$\rho(F*H_T, G*H_T) \leqslant \frac{1}{2\pi}\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|\left(1-\frac{|t|}{T}\right)dt,$$

which, together with (2.2), implies the Esseen theorem.

The condition on the derivative in Theorem 2.1 can be weakened to the Lipschitz condition $\|G\|_{\mathrm{Lip}} \leqslant L$. However, in some problems (where the function G is not necessarily continuous), it is desirable to replace it by the condition of 'average' smoothness. A generalization of this kind was proposed by Fainleib [6] in connection with problems in probabilistic number theory. We present it in the following formulation (see [3], Chap. V, § 1, Theorem 1).

Theorem 2.2. *Let F be a non-decreasing bounded function and let G be a function of bounded variation such that $F(-\infty) = G(-\infty) = 0$. Then for any $T > 0$ and all $b > 1/(2\pi)$*

$$\rho(F,G) \leqslant b\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|dt + bT\sup_x \int_{|u|\leqslant r(b)/T}|G(x+u)-G(x)|\,du, \tag{2.4}$$

where the constant $r(b)$ depends only on b.

We can simplify this inequality by formulating it in terms of the modulus of continuity

$$Q_G(h) = \sup_{x\in\mathbb{R},\,|u|\leqslant h}|G(x+u)-G(x)|, \qquad h \geqslant 0. \tag{2.5}$$

If G is a distribution function, then Q_G is called the concentration function, and in this case it can be estimated from above with the help of the well-known inequality

$$Q_G(h) \leqslant \left(\frac{96}{95}\right)^2 h\int_{-1/h}^{1/h}|g(t)|\,dt, \qquad h > 0$$

(see [3], Chap. III, § 1, Lemma 6). Therefore, Theorem 2.2 implies the following statement.

Corollary 2.3. *Under the assumptions of Theorem 2.2, for any $T>0$*

$$c\rho(F,G)\leqslant\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|dt+Q_G\left(\frac{1}{T}\right), \tag{2.6}$$

where c is an absolute positive constant. Moreover, if G is non-decreasing, then

$$c\rho(F,G)\leqslant\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|dt+\frac{1}{T}\int_{-T}^{T}|g(t)|\,dt. \tag{2.7}$$

An advantage of (2.7) over (2.1) is that its right-hand side does not contain a constant value bounding $|G'|$, but is a functional depending directly on the characteristic function g. In this connection, we note another inequality:

$$\rho(F,G)\leqslant\frac{1}{2\pi}\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|dt+\frac{1}{T}\int_{-T}^{T}(|f(t)|+|g(t)|)\,dt, \tag{2.8}$$

which is valid for arbitrary distribution functions F and G with characteristic functions f and g (up to an absolute multiplicative constant, it is weaker than (2.7)). The inequality (2.8) was obtained by Bentkus and Götze [7] as a corollary to a smoothing inequality of Prawitz (see (18.7)); it has been successfully used in problems on the number of integer points inside multidimensional ellipsoids, as well as in investigations of the asymptotic behaviour of the distribution of quadratic forms in sums of independent random vectors.

3. Lévy distance

The Lévy distance between distribution functions F and G is defined by

$$L(F,G)=\inf\bigl\{h\geqslant 0\colon G(x-h)-h\leqslant F(x)\leqslant G(x+h)+h\ \forall\,x\in\mathbb{R}\bigr\}.$$

It is straightforward to see that $L(F,G)$ is the side length of the largest square lying between the graphs of F and G (with the sides of the square parallel to the coordinate axes).

The Lévy distance turns $\mathscr{F}$ into a metric space with the topology of weak convergence. Namely, $L(F_n,F)\to 0$ as $n\to\infty$ if and only if $\lim_{n\to\infty}F_n(x)=F(x)$ at all points of continuity of F. And therefore this metric is finding more and more applications in problems where one needs to estimate the proximity of distributions in the sense of the weak topology.

The Kolmogorov distance is obviously related to the Lévy distance by the inequalities

$$0\leqslant L(F,G)\leqslant\rho(F,G)\leqslant 1,$$

and hence the former induces a stronger topology in $\mathscr{F}$. On the other hand, if it is known that $|G'(x)|\leqslant L$ (as in Esseen's theorem), then we have the reverse inequality

$$\rho(F,G)\leqslant(1+L)L(F,G).$$

In this case, smoothing inequalities for these distances are equivalent. However, it is generally natural to expect that estimates for the proximity of F and G in the

Lévy metric in terms of the proximity of their characteristic functions should hold under less restrictive smoothness-type conditions or even under none of these at all. Results of this kind were first obtained by Bohman [8], who proved the following statement.

Theorem 3.1. *Let F and G be distribution functions with characteristic functions f and g, respectively. If $|f(t)-g(t)| \leqslant \lambda|t|$ for all $t \in \mathbb{R}$, then for any $x \in \mathbb{R}$ and $h>0$*

$$G(x-h)-\frac{2\lambda}{h} \leqslant F(x) \leqslant G(x+h)+\frac{2\lambda}{h}\,. \tag{3.1}$$

In particular,

$$\frac{1}{2}L^2(F,G) \leqslant \sup_t\left|\frac{f(t)-g(t)}{t}\right|. \tag{3.2}$$

Proof. We give a simple argument to prove this theorem. Without loss of generality we assume that the function $(f(t)-g(t))/t$ is integrable on the whole real line, so that the function

$$\widehat{A}(t)=e^{-itx}\,\frac{f(t)-g(t)}{-it}$$

is the Fourier transform of $A(u)=F(x+u)-G(x+u)$ (for fixed x). Since

$$\int_{-\infty}^{\infty} e^{itu}(1-|u|)_+\,du = 2\frac{1-\cos t}{t^2}\,,$$

the function $\widehat{p}(t)=2(1-\cos t)/t^2$ is the characteristic function for the (probability) triangular density $p(u)=(1-|u|)_+$. It follows that $\widehat{p}_h(t)=\widehat{p}(ht)$ is the characteristic function for the density $p_h(u)=p(u/h)/h$.

By Parseval's identity,

$$I \equiv \int_{-\infty}^{\infty} p_h(u)A(u)\,du = \frac{1}{2\pi}\int_{-\infty}^{\infty}\widehat{p}_h(t)\widehat{A}(t)\,dt.$$

The condition $|f(t)-g(t)| \leqslant \lambda|t|$ implies the estimate $|\widehat{A}(t)| \leqslant \lambda$, and we get that

$$|I| \leqslant \frac{\lambda}{2\pi}\int_{-\infty}^{\infty}\widehat{p}_h(t)\,dt = \frac{\lambda}{2\pi h}\int_{-\infty}^{\infty}\widehat{p}(t)\,dt = \frac{\lambda}{2\pi h}\cdot(2\pi)p(0) = \frac{\lambda}{h}\,.$$

On the other hand, using the monotonicity of F and G along with the fact that p_h has the support $[-h,h]$, we obtain the lower estimate

$$I = \int_{-h}^{h} p_h(u)\big(F(x+u)-G(x+u)\big)\,du \geqslant F(x-h)-G(x+h).$$

The last two estimates yield the inequality $F(x-h) \leqslant G(x+h)+\lambda/h$, which is equivalent to the right-hand inequality in (3.1). Similarly, we obtain the estimate $-I \geqslant G(x-h)-F(x+h)$, which implies the left-hand inequality in (3.1). $\square$

Further generalizations and improvements of Bohman's results were obtained by Zolotarev. In particular, he derived the family of estimates

$$L(F,G) \leqslant C_\gamma\bigg(\sup_{t>0}\,\frac{|f(t)-g(t)|}{t^\gamma}\bigg)^{1/(1+\gamma)}, \qquad \gamma>0,$$

where we can put $C_\gamma = (2/\gamma)(1+\gamma)^2\pi^{-1/(1+\gamma)}$ (see [9] and [10]). However, most popular is the following inequality, proved in [11].

Theorem 3.2. *Let F and G be distribution functions with characteristic functions f and g, respectively. Then for any $T > 1.3$*

$$L(F,G) \leqslant \frac{1}{2\pi}\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right| dt + 2e\frac{\log T}{T}\,. \tag{3.3}$$

It was later shown by Zaitsev [12] that the logarithmic factor in (3.3) is essential. The derivation of (3.3) in [11] was based on the smoothing inequality

$$L(F,G) \leqslant L(F*H, G*H) + \max\{2\varepsilon, 1-H(\varepsilon)+H(-\varepsilon)\},$$

where H can be an arbitrary distribution function and $\varepsilon > 0$ is also arbitrary. This smoothing inequality was applied to convolution powers of the triangular density, and the scale was changed in such a way that H, regarded as a measure, was concentrated on the interval $[-1,1]$.

We give another proof of (a variant of) (3.3) that is based on the smoothing of F and G with the help of a normal distribution function

$$\Phi_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/(2\sigma^2)}\,dy, \qquad x\in\mathbb{R},$$

where the parameter $\sigma > 0$ is chosen depending on T.

We need to estimate from above the expression

$$\mathscr{L}(x,h) = \max\{F(x-h)-G(x+h),\, G(x-h)-F(x+h)\}, \qquad x\in\mathbb{R}, \quad h\geqslant 0,$$

which is related to the Lévy distance through the implication

$$\sup_x \mathscr{L}(x,h) \leqslant b+2h \quad\Longrightarrow\quad L(F,G)\leqslant b+2h, \qquad b\geqslant 0, \quad h\geqslant 0. \tag{3.4}$$

For the convolutions $F_\sigma = F*\Phi_\sigma$ and $G_\sigma = G*\Phi_\sigma$ we consider the deviation

$$I \equiv F_\sigma(x) - G_\sigma(x) = \int_{-\infty}^{\infty}\bigl(F(x-\sigma y)-G(x-\sigma y)\bigr)\,d\Phi(y),$$

where $\Phi = \Phi_1$ is the standard normal distribution function. Using the Kolmogorov distance and the inequality (1.3) for the pair (F_σ, G_σ), we get the uniform estimate

$$|I| \leqslant \rho(F_\sigma, G_\sigma) \leqslant \frac{1}{2\pi}\int_{-\infty}^{\infty}\left|\frac{f(t)-g(t)}{t}\right| e^{-\sigma^2t^2/2}\,dt. \tag{3.5}$$

Let us now estimate the integral I from below by splitting it into two parts $I = I_0 + I_1$, where

$$I_0 = \int_{|y|\leqslant l}\bigl(F(x-\sigma y)-G(x-\sigma y)\bigr)\,d\Phi(y),$$
$$I_1 = \int_{|y|\geqslant l}\bigl(F(x-\sigma y)-G(x-\sigma y)\bigr)\,d\Phi(y)$$

with a parameter $l > 0$. Like I, the second integral can be estimated from above as

$$|I_1| \leqslant 2(1 - \Phi(l)) \equiv \gamma. \tag{3.6}$$

At the same time, using the monotonicity of F and G, we obtain for the first integral the lower estimate

$$I_0 \geqslant \big(F(x - \sigma l) - G(x + \sigma l)\big)(1 - \gamma).$$

Similarly,

$$-I_0 \geqslant \big(G(x - \sigma l) - F(x + \sigma l)\big)(1 - \gamma),$$

so that $|I_0| \geqslant (1-\gamma)\mathscr{L}(x, \sigma l)$. Since $|I_0| \leqslant |I| + |I_1|$, the inequalities (3.5) and (3.6) imply that

$$\sup_x \mathscr{L}(x, \sigma l) \leqslant \frac{1}{1-\gamma}\rho(F_\sigma, G_\sigma) + \frac{\gamma}{1-\gamma}\,. \tag{3.7}$$

Now we estimate the integral in (3.5) by using the elementary inequality

$$\int_T^\infty \frac{1}{t} e^{-\sigma^2 t^2/2}\,dt < \frac{1}{\sigma^2 T^2} e^{-\sigma^2 T^2/2}.$$

Since $|f(t) - g(t)| \leqslant 2$, we have

$$\int_T^\infty \frac{|f(t) - g(t)|}{t} e^{-\sigma^2 t^2/2}\,dt \leqslant \frac{2}{\sigma^2 T^2} e^{-\sigma^2 T^2/2},$$

and hence the right-hand side in (3.5) does not exceed

$$\frac{1}{2\pi}\int_{-T}^T \left|\frac{f(t) - g(t)}{t}\right| dt + \frac{2}{\pi\sigma^2 T^2} e^{-\sigma^2 T^2/2}.$$

Returning to (3.7), we see that

$$\sup_x \mathscr{L}(x, \sigma l) \leqslant \frac{1}{2\pi(1-\gamma)}\int_{-T}^T \left|\frac{f(t) - g(t)}{t}\right| dt + \frac{2}{\pi(1-\gamma)\sigma^2 T^2} e^{-\sigma^2 T^2/2} + \frac{\gamma}{1-\gamma}\,.$$

Thus, according to (3.4) with $h = \sigma l$, if we show that

$$\frac{2}{\pi(1-\gamma)\sigma^2 T^2} e^{-\sigma^2 T^2/2} + \frac{\gamma}{1-\gamma} \leqslant 2\sigma l, \tag{3.8}$$

then we arrive at the estimate

$$L(F, G) \leqslant \frac{1}{2\pi(1-\gamma)}\int_{-T}^T \left|\frac{f(t) - g(t)}{t}\right| dt + 2\sigma l. \tag{3.9}$$

We put $\sigma = \dfrac{1}{T}\sqrt{2\log(1+T)}$ and $l = \sqrt{2\log(1+T)}$. For this choice we have

$$\gamma = 2(1 - \Phi(l)) < e^{-l^2/2} = \frac{1}{1+T}\,, \qquad \frac{\gamma}{1-\gamma} \leqslant \frac{1}{T}\,.$$

For example, if $T \geqslant 1$, then

$$l \geqslant \sqrt{\log 4} > 1.17, \quad \gamma < 2(1 - \Phi(1.17)) < 0.26, \quad (\pi(1-\gamma))^{-1} < 0.44,$$

and thus

$$\begin{aligned} \frac{2}{\pi(1-\gamma)\sigma^2 T^2} e^{-\sigma^2 T^2/2} + \frac{\gamma}{1-\gamma} &= \frac{1}{\pi(1-\gamma)(1+T)\log(1+T)} + \frac{\gamma}{1-\gamma} \\ &< \frac{0.44}{(1+T)\log(1+T)} + \frac{2}{1+T} \\ &< \frac{4\log(1+T)}{1+T} < \frac{4\log(1+T)}{T} = 2\sigma l \end{aligned}$$

(here the next-to-last inequality has to be verified only for $T = 1$).

Then (3.8) is satisfied, and we have the estimate (3.9): for all $T \geqslant 1$

$$L(F,G) \leqslant \frac{1}{\pi} \int_{-T}^{T} \left| \frac{f(t) - g(t)}{t} \right| dt + \frac{4\log(1+T)}{T}\,.$$

For $0 < T < 1$ the last term on the right-hand side is greater than 1, and hence this inequality is valid for all $T > 0$.

We give a direct corollary of this inequality.

Corollary 3.3. *If the characteristic functions f and g coincide on the interval $[0,T]$, then*

$$L(F,G) \leqslant \frac{4\log(1+T)}{T}\,.$$

If we use the property $|f(t) - g(t)| \leqslant 2$, then (3.2) implies the asymptotically weaker estimate $L(F,G) \leqslant 2/\sqrt{T}$.

4. Distance in variation

When studying other, stronger metrics on $\mathscr{F}$, we commonly estimate the proximity of smoothed distributions in terms of the total variation (which is one of the strongest metrics). This is completely justified, for as a rule, smoothed distributions have smooth densities, and the distances between them in different metrics are often of the same order. Thus, we concentrate on one standard estimate of the distance in variation in terms of the characteristic functions.

Theorem 4.1. *For any distribution functions F and G with continuously differentiable characteristic functions f and g,*

$$\|F - G\|_{\mathrm{TV}}^4 \leqslant \int_{-\infty}^{\infty} |f(t) - g(t)|^2\,dt \int_{-\infty}^{\infty} |f'(t) - g'(t)|^2\,dt. \tag{4.1}$$

The estimate (4.1) is sometimes written in the formally weaker form (see, for example, [13])

$$\|F - G\|_{\mathrm{TV}}^2 \leqslant \frac{1}{2} \int_{-\infty}^{\infty} |f(t) - g(t)|^2\,dt + \frac{1}{2} \int_{-\infty}^{\infty} |f'(t) - g'(t)|^2\,dt. \tag{4.2}$$

But we arrive at (4.1) by changing the scale, or, to be precise, applying (4.2) to $F_r(x) = F(x/r)$, $G_r(x) = G(x/r)$ and optimizing the right-hand side of the resulting inequality over $r > 0$.

The inequality (4.1) remains valid for arbitrary functions F and G of bounded variation. The finiteness of the integrals in (4.1) provides the absolute continuity of the function $A = F - G$. Consequently, if G has a density q, then F should have a density p, and so (4.1) turns into

$$\biggl(\int_{-\infty}^{\infty} |p(x) - q(x)|\,dx\biggr)^4 \leqslant \int_{-\infty}^{\infty} |f(t) - g(t)|^2\,dt \int_{-\infty}^{\infty} |f'(t) - g'(t)|^2\,dt.$$

Therefore, Theorem 4.1 has a more general statement, in which, moreover, the condition of continuous differentiability of the Fourier transforms can be weakened to absolute continuity.

First we recall that the total variation $\|A\|_{\mathrm{TV}}$ of a complex-valued function A on the real line is defined as the least upper bound of the sums

$$\sum_{k=1}^{n} |A(x_k) - A(x_{k-1})|$$

over all finite collections $x_0 < x_1 < \cdots < x_n$, and the boundedness of variation means that $\|A\|_{\mathrm{TV}} < \infty$. In this case the limits $A(-\infty) = \lim_{x\to-\infty} A(x)$ and $A(\infty) = \lim_{x\to\infty} A(x)$ are finite, and, without loss of generality, one can always assume that A is right-continuous and $A(-\infty) = 0$. If A is (locally) absolutely continuous, then

$$\|A\|_{\mathrm{TV}} = \int_{-\infty}^{\infty} |A'(x)|\,dx,$$

where A' is the Radon–Nikodym derivative. It is often convenient to identify A with a Borel (complex-valued) measure on $\mathbb{R}$ that is determined by the equality $A((x,y]) = A(y) - A(x)$, $x < y$. Then A' is the Radon–Nikodym derivative of the measure A with respect to the linear Lebesgue measure. The uniqueness of A' is understood to within its values on a zero-measure set (that is, in the space $L^1(\mathbb{R})$).

Theorem 4.2. *Any (locally) absolutely continuous function $a\colon \mathbb{R} \to \mathbb{C}$ such that $\int |a(t)|^2\,dt < \infty$ and $\int |a'(t)|^2\,dt < \infty$ is the Fourier transform of an integrable function $b\colon \mathbb{R} \to \mathbb{C}$ (which is unique in L^1), that is,*

$$a(t) = \int_{-\infty}^{\infty} e^{itx} b(x)\,dx, \qquad t \in \mathbb{R}. \tag{4.3}$$

Furthermore, in this case

$$\biggl(\int_{-\infty}^{\infty} |b(x)|\,dx\biggr)^4 \leqslant \int_{-\infty}^{\infty} |a(t)|^2\,dt \int_{-\infty}^{\infty} |a'(t)|^2\,dt. \tag{4.4}$$

Proof. We prove the theorem using standard arguments. First we assume that a and a' belong to $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ and we introduce the function

$$b(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} a(t)\,dt. \tag{4.5}$$

Let us show that it is integrable on the whole real line.

The assumption $a \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ guarantees that $b \in L^2(\mathbb{R})$, and Plancherel's formula holds:

$$\int_{-\infty}^{\infty} |b(x)|^2\,dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} |a(t)|^2\,dt. \tag{4.6}$$

The conditions $a \in L^1(\mathbb{R})$ and $a' \in L^1(\mathbb{R})$ let us deduce from (4.5) the equality

$$(ix)b(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}a'(t)\,dt. \tag{4.7}$$

Indeed, we note that since a' is integrable, the function a has bounded variation, and thus the limits $a(-\infty)$ and $a(\infty)$ exist and are finite. Moreover, $a(-\infty) = a(\infty) = 0$ because a is integrable. Consequently, choosing an arbitrary $N > 0$, we can integrate by parts:

$$\int_{-N}^{N} e^{-itx}a(t)\,dt = a(t)\frac{e^{-itx}}{-ix}\bigg|_{t=-N}^{t=N} + \frac{1}{ix}\int_{-N}^{N} e^{-itx}a'(t)\,dt, \qquad x \neq 0.$$

Letting N go to ∞, we arrive at (4.7).

The condition $a' \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ allows us to apply Plancherel's formula on the basis of (4.7), and we obtain

$$\int_{-\infty}^{\infty} x^2|b(x)|^2\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |a'(t)|^2\,dt.$$

Combining this with (4.6), we derive the equality

$$\int_{-\infty}^{\infty} |1+ix|^2|b(x)|^2\,dx = \frac{1}{2\pi}\bigg(\int_{-\infty}^{\infty} |a(t)|^2\,dt + \int_{-\infty}^{\infty} |a'(t)|^2\,dt\bigg).$$

Then by the Cauchy–Schwarz inequality, we have

$$\begin{aligned}\bigg(\int_{-\infty}^{\infty} |b(x)|\,dx\bigg)^2 &= \bigg(\int_{-\infty}^{\infty} |1+ix|\,|b(x)|\frac{1}{|1+ix|}\,dx\bigg)^2 \\ &\leqslant \int_{-\infty}^{\infty} |1+ix|^2|b(x)|^2\,dx \int_{-\infty}^{\infty} \frac{1}{|1+ix|^2}\,dx \\ &\leqslant \frac{1}{2}\int_{-\infty}^{\infty} |a(t)|^2\,dt + \frac{1}{2}\int_{-\infty}^{\infty} |a'(t)|^2\,dt.\end{aligned}$$

Applying this to $a_r(x) = a(rx)$ and optimizing over r, we obtain (4.4). Therefore, b is indeed integrable, and hence the inverse Fourier transform can be used on the basis of (4.5), yielding (4.3). Thus, the theorem is proved under the additional assumptions.

Let us now consider the general case, and first reformulate the theorem in operator language. We denote the usual L^p-norm of functions (with respect to Lebesgue measure) by $\|\cdot\|_p$. The basic function space in Theorem 4.2 is the Sobolev space

$W_1^2 = W_1^2(\mathbb{R})$ of all absolutely continuous complex-valued functions $a(t)$ on the real line with finite norm

$$\|a\|_{W_1^2} = \left(\frac{1}{2}\|a\|_2^2 + \frac{1}{2}\|a'\|_2^2\right)^{1/2}.$$

We need the following assertion.

Lemma 4.3. *The linear subspace* $H = \{a \in W_1^2 \colon a \in L^1(\mathbb{R}),\ a' \in L^1(\mathbb{R})\}$ *is dense in* W_1^2.

To see this, we note that every function $a \in H$ can be approximated in the norm of W_1^2 by functions

$$a_\sigma(t) = a(t)\, e^{-\sigma^2 t^2/2}, \qquad \sigma > 0.$$

Obviously, $a_\sigma \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Since $|a_\sigma(t)| \leqslant |a(t)|$, by the Lebesgue dominated convergence theorem we have

$$\int_{-\infty}^{\infty} |a_\sigma(t) - a(t)|^2\, dt \to 0, \qquad \sigma \to 0.$$

Moreover, since $a_\sigma'(t) = a'(t) e^{-\sigma^2 t^2/2} - a(t)\sigma^2 t e^{-\sigma^2 t^2/2}$ (where the equality is understood in the Radon–Nikodym sense), we have

$$|a_\sigma'(t) - a'(t)| \leqslant |a'(t)|(1 - e^{-\sigma^2 t^2/2}) + |a(t)|\sigma^2 |t| e^{-\sigma^2 t^2/2}.$$

Hence, using the estimate $xe^{-x^2/2} \leqslant 1/\sqrt{e}$ ($x \geqslant 0$), we find that

$$|a_\sigma'(t) - a'(t)|^2 \leqslant 2|a'(t)|^2 (e^{-\sigma^2 t^2/2} - 1)^2 + \sigma^2 |a(t)|^2,$$

and again by the Lebesgue theorem,

$$\int_{-\infty}^{\infty} |a_\sigma'(t) - a'(t)|^2\, dt \to 0, \qquad \sigma \to 0.$$

Therefore, $\|a_\sigma - a\|_{W_1^2} \to 0$ as $\sigma \to 0$, and hence H is dense in W_1^2.

Continuing the proof of Theorem 4.2, we denote by $\mathscr{P} b = \widehat{b}$ the Fourier transform acting on functions $b \in L^1(\mathbb{R})$. For functions a in H, we considered in the first step the functions $b = \mathscr{G} a$ determined by (4.5), for which we have the identity (4.6) and the inequality (4.4). Consequently,

$$\|\mathscr{G} a\|_1 \leqslant \|a\|_{W_1^2} \quad \text{and} \quad \|\mathscr{G} a\|_2 \leqslant \frac{1}{\sqrt{2\pi}} \|a\|_2 \leqslant \|a\|_{W_1^2}. \tag{4.8}$$

Moreover, $\mathscr{P}\mathscr{G} a = a$. In particular, $\mathscr{G} \colon H \to L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is a continuous linear operator with respect to the norm $\|b\| = \|b\|_1 + \|b\|_2$ in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. But H is dense in W_1^2, and thus $\mathscr{G}$ can be extended by continuity to the entire space W_1^2, with the inequalities in (4.8) remaining valid. Since the operator $\mathscr{P}$ is continuous in the norm of L^2, the identity $\mathscr{P}\mathscr{G} a = a$ also remains true. Then for all $a \in W_1^2$

$$\|b\|_1 \leqslant \left(\frac{1}{2}\|a\|_2^2 + \frac{1}{2}\|a'\|_2^2\right)^{1/2}, \qquad \mathscr{P} b = a, \quad b = \mathscr{G} a.$$

It remains to make this inequality homogeneous over the space variable, in the same way as this was briefly described in the first step. The proof of Theorem 4.2 is complete. □

Remark 4.4. It is not necessary to use the derivatives in (4.1) if it is known that the distributions F and G have compact support. Indeed, we recall the notation in Theorem 4.2. If b in (4.3) is concentrated on an interval $[-R,R]$ with $R>0$, then by applying (4.6) we obtain

$$\begin{aligned}\biggl(\int_{-\infty}^{\infty}|b(x)|\,dx\biggr)^2&=\biggl(\int_{-R}^{R}|b(x)|\,dx\biggr)^2\\&\leqslant 2R\int_{-R}^{R}|b(x)|^2\,dx=\frac{R}{\pi}\int_{-\infty}^{\infty}|a(t)|^2\,dt.\end{aligned}\tag{4.9}$$

In the case $a=f-g$, $a\in L^2$, the function $A=F-G$ is absolutely continuous and, as a measure, has density $b=A'$, so that $\|A\|_{\mathrm{TV}}=\int_{-\infty}^{\infty}|b(x)|\,dx$. Consequently, applying (4.9) to b, we arrive at the following inequality, which is simpler that (4.1).

Theorem 4.5. *For any distribution functions F and G that are concentrated on an interval $[-R,R]$ and have characteristic functions f and g,*

$$\|F-G\|_{\mathrm{TV}}^2\leqslant\frac{R}{\pi}\int_{-\infty}^{\infty}|f(t)-g(t)|^2\,dt.$$

If the integral on the right-hand side is finite, then the function $A=F-G$ is absolutely continuous. As a measure, it has the density $b=A'$, and thus $\|A\|_{\mathrm{TV}}=\int_{-\infty}^{\infty}|b(x)|\,dx$.

5. Kullback–Leibler divergence

The Kullback–Leibler divergence, also called the relative entropy or the information divergence, is an even stronger distance than the total variation of the difference of distributions. For absolutely continuous probability distributions F and G on the real line with densities p and q, this distance is defined by

$$D(F||G)=\int_{-\infty}^{\infty}p(x)\log\frac{p(x)}{q(x)}\,dx$$

under the assumption that $q(x)=0\Rightarrow p(x)=0$ for almost all x (that is, the distribution F regarded as a measure is absolutely continuous with respect to the measure G). In all the other cases one sets $D(F||G)=\infty$.

In the general case we have $0\leqslant D(F||G)\leqslant\infty$, and $D(F||G)=0$ if and only if $F=G$. However, this functional is not symmetric with respect to (F,G) and thus is not a metric in the space $\mathscr{F}$. Nevertheless, in many problems the quantity $D(F||G)$ serves as a convenient measure of the proximity of F to G. It is related to the distance in variation by the well-known Pinsker(-type) inequality

$$D(F||G)\geqslant\frac12\|F-G\|_{\mathrm{TV}}^2=\frac12\biggl(\int_{-\infty}^{\infty}|p(x)-q(x)|\,dx\biggr)^2.$$

We also note the relation to the classical entropy

$$h(X) = -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx,$$

where X is a random variable with density p. Namely, if Z is a Gaussian random variable with distribution G, and X and Z have equal expectations and variances, then

$$D(F||G) = h(Z) - h(X).$$

The problem of estimating the proximity of F and G in the sense of the Kullback–Leibler divergence in terms of the characteristic functions has been little studied. We present a result for the important case when $G = \Phi$ is the standard normal distribution, that is, has the characteristic function $g(t) = e^{-t^2/2}$. Let

$$g_\alpha(t) = e^{-t^2/2}\biggl(1 + \alpha\frac{(it)^3}{3!}\biggr), \qquad t, \alpha \in \mathbb{R}.$$

The function g_α is the Fourier transform of the density

$$\varphi_\alpha(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\biggl(1 + \alpha\frac{x^3 - 3x}{3!}\biggr)$$

of a 'corrected' Gaussian distribution (note that this density can assume negative values).

As before, $\|u\|_2$ denotes the L^2-norm of a function u:

$$\|u\|_2 = \biggl(\int_{-\infty}^{\infty} |u(t)|^2\,dt\biggr)^{1/2}.$$

Theorem 5.1. *Let F be a probability distribution on the real line with the characteristic function f, and let $\int_{-\infty}^{\infty} |x|^3\,dF(x) < \infty$. Then for any $\alpha \in \mathbb{R}$*

$$D(F||\Phi) \leqslant \alpha^2 + 4\bigl(\|f - g_\alpha\|_2 + \|f''' - g_\alpha'''\|_2\bigr). \tag{5.1}$$

The condition of finiteness of the third absolute moment of F guarantees that the characteristic function f has three continuous derivatives.

The inequality (5.1) was proved in [14] in the study of the entropy variant of the central limit theorem. For $\alpha = 0$ it is simpler:

$$D(F||\Phi) \leqslant 4\bigl(\|f - g\|_2 + \|f''' - g'''\|_2\bigr)$$

and can be regarded as a complete analogue of Theorem 4.1. However, taking other values of α leads to more accurate estimates for $D(F||\Phi)$.

6. Lévy–Prokhorov distance

Now we return to weak probability metrics (that is, metrics giving weak convergence of probability measures).

A natural modification of the Lévy distance for Borel probability measures F and G on an arbitrary metric space (M, d) is the Levy–Prokhorov distance $\pi(F, G)$, which was introduced in [15] for metrization of weak convergence. It is defined as the greatest lower bound of values $h \geqslant 0$ such that

$$F(A) \leqslant G(A_h) + h \tag{6.1}$$

for all Borel sets $A \subset M$. Here A_h denotes the open h-neighbourhood of A in the metric d, that is,

$$A_h = \{x \in M\colon \exists\, y \in M,\ d(x, y) < h\}.$$

Though this definition is formally not symmetric, the dual inequality $G(A) \leqslant F(A_h) + h$ easily follows from (6.1).

In the case $M = \mathbb{R}$, if we restricted ourselves in (6.1) only to half-axes A, we would return to the Lévy distance. Thus, the Lévy–Prokhorov distance is stronger: $L(F, G) \leqslant \pi(F, G)$ (in the case of the real line we identify probability measures F with the associated distribution functions $x \mapsto F(-\infty, x])$. Nevertheless, these metrics generate the same topology of weak convergence in $\mathscr{F}$. In the case of the Euclidean space $M = \mathbb{R}^k$, another generalization of the Levý metric is the distance defined by (6.1) for the class of all half-spaces; it is sometimes called the Tsirelson distance.

Yurinskii [16] proposed a variant of estimation of the proximity of probability distributions on $\mathbb{R}^k$ in the Lévy–Prokhorov metric in terms of the proximity of the corresponding characteristic functions under the additional moment conditions

$$\int_{\mathbb{R}^k} \|x\|^{[k/2]+1}\, dF(x) < \infty \quad \text{and} \int_{\mathbb{R}^k} \|x\|^{[k/2]+1}\, dG(x) < \infty \tag{6.2}$$

and under the assumption that G has a density q such that

$$\int_{\mathbb{R}^k} |q(x+h) - q(x)|\, dx \leqslant \lambda \|h\| \tag{6.3}$$

(a smoothness-type condition by analogy with Theorem 2.1). Under these conditions, he obtained the smoothing inequality

$$\pi(F, G) \leqslant c_1 \pi(F * H, G * H) + c_2(1+\lambda) \int_{\mathbb{R}^k} \|x\|\, dH(x),$$

valid for any probability measure H on $\mathbb{R}^k$, with constants c_1 and c_2 depending only on the dimension k. The Lévy–Prokhorov distance between smoothed distributions can be estimated in variation (Theorem 4.1 with $k = 1$), which finally yields the following result [16]. We formulate it in the one-dimensional case, taking into account a remark of Abramov on a possible weakening of the conditions (6.2) in terms of pseudomoments.

Theorem 6.1. *Let F and G be distribution functions with characteristic functions f and g, respectively, and let*

$$\int_{-\infty}^{\infty} |x|\, |F - G|(dx) < \infty. \tag{6.4}$$

If (6.3) is satisfied, then for any $T > 0$ and some absolute constant $c > 0$

$$c\pi(F, G) \leqslant \biggl(\int_{-T}^{T} \bigl(|f(t) - g(t)|^2 + |(f(t) - g(t))'|^2\bigr)\, dt\biggr)^{1/2} + \frac{1+\lambda}{T}\,. \tag{6.5}$$

Here $|F-G|$ is the variation of the function $F-G$, regarded as a finite positive measure on the real line. We note that the continuous differentiability of $f-g$ is guaranteed by the moment condition (6.4).

Although the first term on the right-hand side of (6.5) is in general much larger than the analogous integral term in Esseen's inequality, the main feature of (6.5) is the decrease of the second term (which holds due to (6.3)). Thus, in applications related to the convergence rate in the central limit theorem (when $G=\Phi$), Yurinskii's inequality implies the correct asymptotic behaviour and thereby essentially strengthens the assertion about normal approximation for sums of independent variables (compared to the statement about normal approximation in the sense of the Kolmogorov distance).

In the general case, that is, without additional smoothness-type assumptions, estimates of the proximity $\pi(F,G)$ in terms of the characteristic functions were studied by Abramov [17] and Zaitsev [18], [12] (see also [19]). We give a result obtained in [18] and [12].

Theorem 6.2. *Let F and G be distribution functions satisfying* (6.2) *with $k=1$, and let f and g be their characteristic functions. Then for any $T>e$*

$$c\pi(F,G)\leqslant I(F,G)+\frac{\log T}{T}\,, \tag{6.6}$$

where $c>0$ is an absolute constant and

$$I(F,G)=\biggl(\int_{-T}^{T}|f(t)-g(t)|^2\,dt\int_{-T}^{T}\biggl(\frac{\log T}{T^2}|f(t)-g(t)|^2+|f'(t)-g'(t)|^2\biggr)\,dt\biggr)^{1/4}.$$

The presence of the term $(\log T)/T$ on the right-hand side of (6.6) makes this inequality closer to the Zolotarev inequality for the Lévy distance, whereas the functional $I(F,G)$ reminds us of the estimate (4.1) for the total variation (though with the essential difference that the integrals are now taken over the finite interval $[-T,T]$). The strengthening of the Zolotarev inequality implies, for example, the following statement, which is an improvement of Corollary 3.3 in terms of the Lévy–Prokhorov distance.

Corollary 6.3. *If the characteristic functions f and g coincide on an interval $[0,T]$ with $T>e$, then*

$$\pi(F,G)\leqslant c\frac{\log T}{T}\,,$$

where c is an absolute constant.

7. Distance in the L^p-metric

Another important distance in the space $\mathscr{F}$ is the distance in the metric of the space $L^p(\mathbb{R})$:

$$\|F-G\|_p=\biggl(\int_{-\infty}^{\infty}|F(x)-G(x)|^p\,dx\biggr)^{1/p},\qquad 1\leqslant p<\infty.$$

Unlike the Lévy distance and the Lévy–Prokhorov distance, it is homogeneous of order $1/p$ with respect to the space variable: if random variables X and Y have distribution functions F and G, then for the distribution functions $F_r(x) = F(x/r)$ and $G_r(x) = G(x/r)$ of the random variables rX and rY $(r > 0)$ we have

$$\|F_r - G_r\|_p = r^{1/p}\|F - G\|_p.$$

The function $p \mapsto \|F - G\|_p^p$ is non-increasing, so $\|F - G\|_p \leqslant \|F - G\|_1^{1/p}$. Thus, for finiteness of the L^p-distance, it is sufficient to assume that the absolute moments $\mathsf{E}|X|$ and $\mathsf{E}|Y|$ are finite. Moreover, if $\mathsf{E}|X|^\varepsilon < \infty$ and $\mathsf{E}|Y|^\varepsilon < \infty$ for some $\varepsilon > 0$, then, as is easily verified,

$$\lim_{p\to\infty} \|F - G\|_p = \|F - G\|_\infty = \rho(F, G).$$

We note another elementary relation:

$$L(F, G) \leqslant \|F - G\|_p^{p/(p+1)}. \tag{7.1}$$

Indeed, if $L(F, G) > h \geqslant 0$, then from the definition of the Lévy distance it follows immediately that there exists a point $x_0 \in \mathbb{R}$ such that $G(x_0 - h) - h > F(x_0)$ or $F(x_0) > G(x_0 + h) + h$. For definiteness let the second inequality hold. Then by the monotonicity of F and G,

$$\|F - G\|_p^p \geqslant \int_{x_0}^{x_0+h} |F(x) - G(x)|^p \, dx \geqslant (F(x_0) - G(x_0 + h))^p h \geqslant h^{p+1}.$$

This proves (7.1).

It follows that the topology generated by the L^p-distance is stronger than the topology of weak convergence in $\mathscr{F}$. Nevertheless, convergence in the L^p-metric on any subspace of probability distributions with bounded absolute moments of order $\alpha > 1$, that is, under the condition $\int_{-\infty}^{\infty} |x|^\alpha \, dF(x) \leqslant M$ with a fixed parameter M, is equivalent to weak convergence.

For $p \geqslant 2$, one can estimate the distance $\|F - G\|_p$ in terms of the corresponding characteristic functions by using the classical Hausdorff–Young inequality

$$\|\widehat{a}\|_p \leqslant \|a\|_q, \qquad q = \frac{p}{p-1}, \tag{7.2}$$

which is valid for any integrable complex-valued function a on the real line. Here

$$\widehat{a}(x) = \int_{-\infty}^{\infty} e^{2\pi i t x} a(t) \, dt, \qquad x \in \mathbb{R},$$

is the Fourier transform of the function a with a modified scale.

According to (1.2), the difference $F - G$ is the Fourier transform of the function $a(t) = \dfrac{1}{2\pi} \dfrac{f(t) - g(t)}{-it}$ in the standard sense when $a(t)$ is integrable. Hence it is possible to apply (7.2), and then we obtain the following estimate (in which the assumption of integrability can easily be dropped).

Theorem 7.1. *For any distribution functions F and G with characteristic functions f and g, respectively,*

$$\|F - G\|_p \leqslant \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\frac{f(t) - g(t)}{t}\right|^q dt\right)^{1/q}, \qquad q = \frac{p}{p-1}, \tag{7.3}$$

for $p \geqslant 2$.

This inequality remains true for arbitrary functions F and G of bounded variation such that $F(-\infty) = G(-\infty)$ and $F(\infty) = G(\infty)$ (in this case we can integrate by parts in (1.1) to obtain (1.2)). If we take the limit in (7.3) as $p \to \infty$, then we return to the estimate (1.3) for the Kolmogorov distance.

For $2 < p < \infty$ the constant $1/(2\pi)$ in (7.3) can be made better if we use the improved Hausdorff–Young inequality (see [20]). On the other hand, for $p = 2$ this inequality turns into the following equality in view of Plancherel's theorem:

$$\|F - G\|_2^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{f(t) - g(t)}{t} \right|^2 dt. \tag{7.4}$$

Using this equality in (7.1), we obtain another well-known estimate for the Lévy distance:

$$L(F, G) \leqslant \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \frac{f(t) - g(t)}{t} \right|^2 dt \right)^{1/3}.$$

However, this estimate can hardly compete with Zolotarev's inequality.

Since $|f(t) - g(t)| \leqslant 2$, the proximity of the characteristic functions f and g on a large interval suffices for the proximity of F and G in the L^p-metric. Indeed,

$$\int_{|t| \geqslant T} \left| \frac{f(t) - g(t)}{t} \right|^q dt \leqslant 2^{q+1} \int_T^{\infty} \frac{1}{t^q}\, dt = \frac{2^{q+1}}{q-1} \frac{1}{T^{q-1}}\,.$$

Using the relations

$$\left(\frac{2^{q+1}}{(q-1)T^{q-1}} \right)^{1/q} = 2^{1+1/q} (p-1)^{(p-1)/p} \frac{1}{T^{1/p}} \leqslant \frac{4(p-1)}{T^{1/p}} \quad \text{and} \quad \left(\frac{1}{2\pi} \right)^{1/q} < \frac{1}{2}\,,$$

we obtain the following statement from (7.3).

Corollary 7.2. *Let F and G be distribution functions with characteristic functions f and g. For $p \geqslant 2$ and for any $T > 0$*

$$\|F - G\|_p \leqslant \frac{1}{2} \left(\int_{-T}^{T} \left| \frac{f(t) - g(t)}{t} \right|^q dt \right)^{1/q} + \frac{4(p-1)}{T^{1/p}}\,. \tag{7.5}$$

In particular, if f and g coincide on an interval $[0, T]$, then

$$\|F - G\|_p \leqslant \frac{4(p-1)}{T^{1/p}}\,.$$

8. Distance in the L^1-metric

For $1 \leqslant p < 2$ the inequality (7.3) no longer holds, and other approaches are needed for estimating the proximity of distributions in the L^p-metric in terms of the characteristic functions. Let us consider the most interesting case $p = 1$, when we deal with the mean distance in the space $\mathscr{F}$. In view of the relation between the distance in the L^1-metric and the Kantorovich distance and other transport metrics, we use another standard notation:

$$W_1(F, G) = \|F - G\|_1 = \int_{-\infty}^{\infty} |F(x) - G(x)|\, dx, \qquad F, G \in \mathscr{F}.$$

Recall that, according to (7.1) with $p = 1$, the relation $L^2(F, G) \leqslant W_1(F, G)$ holds.

Let f and g be the characteristic functions for distribution functions F and G. If the quantity $W_1(F,G)$ is finite, then $f-g$ is continuously differentiable, and $a(t)=(f(t)-g(t))/(-it)$ is the Fourier transform of $b=F-G$. Thus, we can use Theorem 4.2, which yields the estimate

$$W_1^2(F,G) \leqslant \frac{1}{2}\int_{-\infty}^{\infty}\left|\frac{f(t)-g(t)}{t}\right|^2 dt + \frac{1}{2}\int_{-\infty}^{\infty}\left|\frac{d}{dt}\,\frac{f(t)-g(t)}{t}\right|^2 dt.$$

Just as done before Corollary 7.2, here we can narrow the integration interval to $[-T,T]$, and then in the best case we obtain an estimate of type

$$W_1(F,G) \leqslant \left(\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|^2 dt\right)^{1/2} + \left(\int_{-T}^{T}\left|\frac{d}{dt}\,\frac{f(t)-g(t)}{t}\right|^2 dt\right)^{1/2} + \frac{c}{\sqrt{T}} \tag{8.1}$$

by analogy with the inequality (7.5) for $p=2$ (and with an additional integral containing the derivatives of the characteristic functions).

To replace the last term $c/\sqrt{T}$ by c/T (which would agree with the power of T in (7.5) for $p=1$), Esseen proposed the use of smoothing by means of a special finite signed measure H_T with parameter $T>0$ on the real line with Fourier–Stieltjes transform

$$h_T(t) = \begin{cases} 1, & |t| \leqslant \dfrac{T}{2}\,, \\ 0, & |t| \geqslant T, \\ \dfrac{2(T-|t|)}{T}\,, & \dfrac{T}{2} \leqslant |t| \leqslant T. \end{cases}$$

Note that $0 \leqslant h_T(t) \leqslant 1$ and $|h_T'(t)| \leqslant 2/T$ ($|t| \neq T, T/2$). In terms of the characteristic functions $v_T(t) = (1-|t|/T)_+$, we can write $h_T = 2v_T - v_{T/2}$. It follows immediately that $\|H_T\|_{\mathrm{TV}} \leqslant 3$.

From this point on, we mainly follow the exposition of the monograph [13], though with some modifications. We denote the class of all complex-valued functions $A=A(x)$ with bounded variation on the real line by V, and the class of their Fourier–Stieltjes transforms $a=a(t)$ by $\widetilde{V}$. Let $\|a\|_{\mathrm{tv}} = \|A\|_{\mathrm{TV}}$.

The class V is closed under convolution, and hence $\widetilde{V}$ is closed under multiplication and is an algebra, in which this norm has the properties

$$\|a+b\|_{\mathrm{tv}} \leqslant \|a\|_{\mathrm{tv}} + \|b\|_{\mathrm{tv}} \quad \text{and} \quad \|ab\|_{\mathrm{tv}} \leqslant \|a\|_{\mathrm{tv}}\|b\|_{\mathrm{tv}}$$

for all $a,b \in \widetilde{V}$. The variation norm does not change if we change the scale, and thus for $a_r(t)=a(rt)$ we also have $\|a_r\|_{\mathrm{tv}} = \|a\|_{\mathrm{tv}}$.

Any characteristic function has norm $\|a\|_{\mathrm{tv}}=1$, and in particular, $\|1\|_{\mathrm{tv}}=1$. In the general case, when the function a is absolutely continuous, the conditions $\int |a(t)|^2\,dt < \infty$ and $\int |a'(t)|^2\,dt < \infty$ guarantee that $a \in \widetilde{V}$, and by Theorem 4.2

$$\|a\|_{\mathrm{tv}} \leqslant \left(\int_{-\infty}^{\infty} |a(t)|^2\,dt \int_{-\infty}^{\infty} |a'(t)|^2\,dt\right)^{1/4}. \tag{8.2}$$

Example 8.1. Let us consider the function

$$u_T(t) = \begin{cases} \dfrac{4t}{-iT^2}\,, & |t| \leqslant \dfrac{T}{2}\,, \\ \dfrac{1}{-it}\,, & |t| \geqslant \dfrac{T}{2}\,. \end{cases}$$

To estimate its norm we can use the identity $u_T(t) = (2/T)u_2(2t/T)$, which reduces the problem to the case $T = 2$. Applying (8.2), we immediately obtain the estimate $\|u_2\|_{\mathrm{tv}} \leqslant \sqrt{8/3}$, and thus

$$\|u_T\|_{\mathrm{tv}} \leqslant \frac{c}{T}\,, \qquad c = 2\sqrt{\frac{8}{3}}\,. \tag{8.3}$$

It should be noted that the problem of minimizing $\|u\|_{\mathrm{tv}}$ in the class of all functions $u \in \widetilde{V}$ such that $u(t) = 1/(-it)$ for $|t| \geqslant 1$ was studied by Beurling (see [4]). It turned out that the minimum value is $\pi/2 = 1.57\ldots$ (the example of u_2 gives the value $\sqrt{8/3} = 1.63\ldots$, which is slightly worse).

Now let A be a given function of bounded variation with Fourier–Stieltjes transform

$$a(t) = \int_{-\infty}^{\infty} e^{itx}\, dA(x).$$

If A is integrable, then $A(-\infty) = A(\infty) = 0$ and the function a is absolutely continuous (locally). Moreover, we can integrate by parts:

$$b(t) \equiv \frac{a(t)}{-it} = \int_{-\infty}^{\infty} e^{itx}\, A(x)\, dx = \int_{-\infty}^{\infty} e^{itx}\, dB(x), \qquad B(x) = \int_{-\infty}^{x} A(y)\, dy.$$

Hence, b belongs to $\widetilde{V}$ and has the norm

$$\|b\|_{\mathrm{tv}} = \|B\|_{\mathrm{TV}} = \int_{-\infty}^{\infty} |A(x)|\, dx.$$

On the other hand, if we use the representation $b = bh_T + b\cdot(1-h_T)$, then we get by the triangle inequality that

$$\|b\|_{\mathrm{tv}} \leqslant \|bh_T\|_{\mathrm{tv}} + \|b\cdot(1-h_T)\|_{\mathrm{tv}}. \tag{8.4}$$

Since $1 - h_T(t) = 0$ for $|t| \leqslant T/2$, the equality

$$b(t)(1-h_T(t)) = a(t)u(t)(1-h_T(t))$$

holds for any function $u \in \widetilde{V}$ such that $u(t) = 1/(-it)$ for $|t| \geqslant T/2$. Moreover, in this case

$$\|b\cdot(1-h_T)\|_{\mathrm{tv}} \leqslant \|a\|_{\mathrm{tv}}\|u\|_{\mathrm{tv}}\|1-h_T\|_{\mathrm{tv}}.$$

Here $\|a\|_{\mathrm{tv}} = \|A\|_{\mathrm{TV}}$ and $\|1-h_T\|_{\mathrm{tv}} \leqslant 1 + \|h_T\|_{\mathrm{tv}} \leqslant 4$, and hence

$$\|b\cdot(1-h_T)\|_{\mathrm{tv}} \leqslant 4\|A\|_{\mathrm{TV}}\|u\|_{\mathrm{tv}}.$$

Taking the function u_T in Example 8.1 as u, we get from (8.4) the smoothing-type inequality

$$\int_{-\infty}^{\infty} |A(x)|\,dx \leqslant \|bh_T\|_{\mathrm{tv}} + \frac{4c\|A\|_{\mathrm{TV}}}{T}\,, \qquad c = 2\sqrt{\frac{8}{3}}\,. \tag{8.5}$$

Let us consider the first term on the right-hand side of this inequality. Applying (8.2) and using the fact that $h_T = 0$ outside the interval $(-T,T)$, we arrive at the inequality

$$\|bh_T\|_{\mathrm{tv}}^2 \leqslant \frac{1}{2}\int_{-T}^{T} |b(t)h_T(t)|^2\,dt + \frac{1}{2}\int_{-T}^{T} |(b(t)h_T(t))'|^2\,dt. \tag{8.6}$$

Here the first integral does not exceed $\varepsilon = \int_{-T}^{T} |b(t)|^2\,dt$. In addition, the identity $(bh_T)' = b'h_T + bh_T'$ (in the sense of Radon–Nikodym) for $|t| < T$ implies the inequality

$$|(b(t)h_T(t))'|^2 \leqslant 2|b'(t)|^2 + \frac{8|b(t)|^2}{T^2}\,,$$

which holds almost everywhere. Consequently, putting $\delta = \int_{-T}^{T} |b'(t)|^2\,dt$, we can estimate the second integral in (8.6) by the quantity $2\delta + (8/T^2)\varepsilon$ and finally obtain

$$\|bh_T\|_{\mathrm{tv}}^2 \leqslant \biggl(\frac{1}{2} + \frac{4}{T^2}\biggr)\varepsilon + \delta.$$

The last inequality together with (8.5) yields the following theorem of Esseen [21].

Theorem 8.2. *Let A be a complex-valued integrable function of bounded variation with Fourier–Stieltjes transform a. Then for all $T > 0$*

$$\int_{-\infty}^{\infty} |A(x)|\,dx \leqslant c_T\biggl(\int_{-T}^{T}\biggl|\frac{a(t)}{t}\biggr|^2\,dt\biggr)^{1/2} + \biggl(\int_{-T}^{T}\biggl|\frac{d}{dt}\,\frac{a(t)}{t}\biggr|^2\,dt\biggr)^{1/2} + c\frac{\|A\|_{\mathrm{TV}}}{T}\,, \tag{8.7}$$

where $c_T = (1/2 + 4/T^2)^{1/2}$ and c is a constant. In particular, (8.7) holds with $c = 16\sqrt{2/3} < 13.07$.

If $A = F - G$, where F and G are distribution functions, then $\|A\|_{\mathrm{TV}} \leqslant 2$, and we obtain the following improvement of (8.1).

Corollary 8.3. *Let F and G be distribution functions with characteristic functions f and g. If $W_1(F,G) < \infty$, then for all $T \geqslant 3$*

$$W_1(F,G) \leqslant \biggl(\int_{-T}^{T}\biggl|\frac{f(t)-g(t)}{t}\biggr|^2\,dt\biggr)^{1/2} + \biggl(\int_{-T}^{T}\biggl|\frac{d}{dt}\,\frac{f(t)-g(t)}{t}\biggr|^2\,dt\biggr)^{1/2} + \frac{c}{T}\,, \tag{8.8}$$

where c is a constant. In particular, if f and g coincide on $[0,T]$, then $W_1(F,G) \leqslant c/T$.

By (8.7), we can take $c = 32\sqrt{2/3}$ in the last inequality. This constant can be improved on the basis of the indicated result of Beurling. Esseen [4] showed that $W_1(F,G) \leqslant \pi/T$ in the case when the characteristic functions f and g coincide on the interval $[0,T]$.

Finally, to draw attention to the relationship between Theorems 8.2 and 4.2, we reformulate (8.7) and observe that $a(t)/(-it)$ is the Fourier transform of A. Upon changing the notation, we obtain the following assertion.

Corollary 8.4. *For any integrable function $b\colon \mathbb{R} \to \mathbb{C}$ of bounded variation, its Fourier transform $\widehat{b}$ is a (locally) absolutely continuous function, and for all $T > 0$*

$$\int_{-\infty}^{\infty} |b(x)|\,dx \leqslant \biggl(\frac{1}{2} + \frac{4}{T^2}\biggr)^{1/2} \biggl(\int_{-T}^{T} |\widehat{b}(t)|^2\,dt\biggr)^{1/2} + \biggl(\int_{-T}^{T} \biggl|\frac{d}{dt}\widehat{b}(t)\biggr|^2\,dt\biggr)^{1/2} + \frac{14}{T}\|b\|_{\mathrm{TV}}.$$

Letting T tend to infinity, we derive an inequality that is equivalent to (4.4) up to an absolute multiplicative constant.

9. Ideal Zolotarev metrics

In the mid-1970s, in connection with problems on the rate of convergence in the central limit theorem, Zolotarev introduced the so-called ideal metrics in the space of probability distributions on $\mathbb{R}^k$. Among these metrics the following are the most important (see, for example, [22], [23]). For simplicity we consider only the one-dimensional case.

Fix an integer $s \geqslant 0$. For probability distributions F and G on the real line that have finite absolute moments of order s, we put

$$\zeta_s(F,G) = \sup\biggl|\int_{-\infty}^{\infty} u\,dF - \int_{-\infty}^{\infty} u\,dG\biggr|, \tag{9.1}$$

where the supremum is taken over all functions $u\colon \mathbb{R} \to \mathbb{R}$ having a derivative of order $s-1$ satisfying the Lipschitz condition:

$$|u^{(s-1)}(x) - u^{(s-1)}(y)| \leqslant |x-y|, \qquad x, y \in \mathbb{R}$$

(it is sufficient to consider s times differentiable functions u such that $|u^{(s)}| \leqslant 1$).

In the case $s = 0$ we obtain the distance in variation $\zeta_0(F,G) = \|F-G\|_{\mathrm{TV}}$.

In the case $s = 1$ it is easy to see that we return to the metric in L^1:

$$\zeta_1(F,G) = \sup_{\|u\|_{\mathrm{Lip}} \leqslant 1} \biggl|\int_{-\infty}^{\infty} u\,dF - \int_{-\infty}^{\infty} u\,dG\biggr| = \int_{-\infty}^{\infty} |F(x) - G(x)|\,dx.$$

In the case $s = 2$ a similar formula holds:

$$\zeta_2(F,G) = \int_{-\infty}^{\infty} \biggl|\int_{-\infty}^{x} (F(y) - G(y))\,dy\biggr|\,dx.$$

In the general case we let $b_1 = F - G$ and obtain the recurrence relation

$$\zeta_s(F,G) = \int_{-\infty}^{\infty} |b_s(x)|\,dx, \qquad b_s(x) = \int_{-\infty}^{x} b_{s-1}(y)\,dy, \tag{9.2}$$

which yields the representation

$$\zeta_s(F,G) = \frac{1}{(s-2)!}\int_{-\infty}^{\infty}\left|\int_{-\infty}^{x}(F(y)-G(y))\,(x-y)^{s-2}\,dy\right|dx, \qquad s \geqslant 2.$$

Note that for the finiteness of $\zeta_s(F,G)$ it suffices that F and G have coinciding moments of orders up to $s-1$ inclusive:

$$\int_{-\infty}^{\infty} x^p\,dF(x) = \int_{-\infty}^{\infty} x^p\,dG(x), \qquad p = 1,\dots,s-1, \tag{9.3}$$

and also have finite absolute moments $\int_{-\infty}^{\infty}|x|^s\,dF(x)$ and $\int_{-\infty}^{\infty}|x|^s\,dG(x)$ (which was assumed from the beginning).

The distance ζ_s is homogeneous of order s with respect to the space variable: if random variables X and Y have distribution functions F and G, then the distribution functions F_r and G_r of the random variables rX and rY $(r > 0)$ satisfy

$$\zeta_s(F_r, G_r) = r^s \zeta_s(F,G).$$

These metrics are related; in particular,

$$\zeta_1 \leqslant C_s \zeta_{1+s}^{1/(1+s)}, \qquad \zeta_2^{1/2} \leqslant C_s' \zeta_{2+s}^{1/(2+s)},$$

where the constants depend only on s (see [9], Theorem 3). For example, $\zeta_1^2 \leqslant 8\zeta_2$ (see [23]).

Using ζ_2, we can estimate the Lévy–Prokhorov distance by

$$\pi^3(F,G) \leqslant c\zeta_2(F,G), \qquad F,G \in \mathscr{F},$$

with some constant c (see [24]).

The Zolotarev metrics can be defined similarly also for non-integers $s = m + \alpha$ (m an integer, $0 < \alpha < 1$). In this case the supremum in (9.1) is taken over all m times differentiable functions u having derivatives of order m that satisfy the Lipschitz condition with exponent α:

$$|u^{(m)}(x) - u^{(m)}(y)| \leqslant |x-y|^\alpha, \qquad x,y \in \mathbb{R}.$$

By (9.1), the definition of $\zeta_s(F,G)$ can be extended to any functions F and G of bounded variation. For finiteness of $\zeta_1(F,G)$, it should be assumed that the integral $\int_{-\infty}^{\infty}|F(x)-G(x)|\,dx$ is finite, which, in turn, is guaranteed by the condition

$$\int_{-\infty}^{\infty} |x|\,\bigl|d\bigl(F(x)-G(x)\bigr)\bigr| < \infty,$$

and then we necessarily have $F(-\infty) = G(-\infty)$ and $F(\infty) = G(\infty)$.

In this more general case, we pass to the problem of estimating the Zolotarev distance in terms of the Fourier–Stieltjes transforms

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x) \quad \text{and} \quad g(t) = \int_{-\infty}^{\infty} e^{itx}\, dG(x), \qquad t \in \mathbb{R}.$$

To estimate $\zeta_1(F,G)$ one can use Theorem 8.2 or Corollary 8.4: the inequality

$$\zeta_1(F,G) \leqslant \bigg(\int_{-T}^{T} \bigg|\frac{f(t)-g(t)}{t}\bigg|^2 dt\bigg)^{1/2} + \bigg(\int_{-T}^{T} \bigg|\frac{d}{dt}\,\frac{f(t)-g(t)}{t}\bigg|^2 dt\bigg)^{1/2} + \frac{14}{T}\|F-G\|_{\mathrm{TV}}$$

holds for all $T \geqslant 3$.

If the function $b_2(x) = \int_{-\infty}^{x} (F(y)-G(y))\, dy$ is integrable, then we can integrate by parts on the right-hand side of the equality

$$\frac{f(t)-g(t)}{-it} = \int_{-\infty}^{\infty} e^{itx}(F(x)-G(x))\, dx = \int_{-\infty}^{\infty} e^{itx}\, db_2(x),$$

which implies that

$$\frac{f(t)-g(t)}{(-it)^2} = \int_{-\infty}^{\infty} e^{itx} b_2(x)\, dx.$$

Applying Corollary 8.4 to $b = b_2$ and taking the equality $\|b_2\|_{\mathrm{TV}} = \zeta_1(F,G)$ into account, we now get that

$$\zeta_2(F,G) \leqslant \bigg(\int_{-T}^{T} \bigg|\frac{f(t)-g(t)}{t^2}\bigg|^2 dt\bigg)^{1/2} + \bigg(\int_{-T}^{T} \bigg|\frac{d}{dt}\,\frac{f(t)-g(t)}{t^2}\bigg|^2 dt\bigg)^{1/2} + \frac{14}{T}\zeta_1(F,G).$$

Note that the integrability of b_2 is guaranteed by the conditions

$$\int_{-\infty}^{\infty} x\, dF(x) = \int_{-\infty}^{\infty} x\, dG(x), \qquad \int_{-\infty}^{\infty} |x|^2\, |d(F(x)-G(x))| < \infty,$$

where, as usual, $|d(F(x)-G(x))|$ denotes the variation of $F-G$ (which is a finite positive measure on the real line).

Proceeding on the basis of the recurrence formula (9.2), we arrive at similar estimates for arbitrary integers $s \geqslant 1$.

Theorem 9.1. *Let F and G be functions of bounded variation that have identical moments of all orders up to $s-1$ inclusive* (*the condition* (9.3)) *and are such that*

$$\int_{-\infty}^{\infty} |x|^s\, \big|d\big(F(x)-G(x)\big)\big| < \infty.$$

Then for all $T \geqslant 3$

$$\zeta_s(F,G) \leqslant \bigg(\int_{-T}^{T} \bigg|\frac{f(t)-g(t)}{t^s}\bigg|^2 dt\bigg)^{1/2} + \bigg(\int_{-T}^{T} \bigg|\frac{d}{dt}\,\frac{f(t)-g(t)}{t^s}\bigg|^2 dt\bigg)^{1/2} + \frac{c}{T}\,\zeta_{s-1}(F,G)$$

with some absolute constant $c > 0$ (*one can take $c = 14$*).

In the class of distribution functions, this inequality was proved by Zolotarev [25] (up to an absolute constant); it is also given (without proof) in [23], p. 80.

10. Transport metrics

Let (M,d) be a complete separable metric space, and denote by $\mathscr{F}_p(M)$ the space of all (Borel) probability measures F on M that have a finite moment of order $p\geqslant 1$, that is, satisfy the condition

$$\int \bigl(d(x,x_0)\bigr)^p\,dF(x)<\infty$$

for some (and hence for all) $x_0\in M$. We put

$$W_p(F,G)=\inf\biggl(\iint \bigl(d(x,y)\bigr)^p\,d\mu(x,y)\biggr)^{1/p},$$

where the infinum is taken over all probability measures μ on $M\times M$ with marginal distributions F and G, that is, over measures such that

$$\mu(A\times M)=F(A),\qquad \mu(M\times A)=G(A)$$

for all Borel sets $A\subset M$. The functional W_p turns $\mathscr{F}_p(M)$ into a metric space.

According to Vershik's historical study [26], the distance W_1 was introduced by Kantorovich in the late 1930s. The latter also considered more general functionals of the form

$$W=\inf\iint c(x,y)\,d\mu(x,y),$$

with the following interpretation: if the cost of transporting a 'particle' from point x to point y is $c(x,y)$, then the cost of optimal transportation of a 'mass' F to G is equal to W (see [27] and [28]). Therefore, the distances W_p are also called transport distances or minimal distances. A detailed discussion of many important properties and applications of these metrics can be found in [29]–[32]. Here we mention some of them.

As follows directly from the definition, the function $p\mapsto W_p(F,G)$ is non-decreasing, and hence the metric W_p becomes stronger with growing p. The Kantorovich distance is related to the Lévy–Prokhorov distance by the inequality

$$\pi(F,G)\leqslant \bigl(W_p(F,G)\bigr)^{p/(p+1)}. \tag{10.1}$$

This is an analogue of the relation (7.1) between the Lévy metric and the metric in the space L^p. We give a similar proof. Assume that $F(A)\geqslant G(A_h)+h$ for some $h>0$ and some Borel set A in M. Then for any probability measure μ on $M\times M$ with marginal distributions F and G we have

$$\begin{aligned}\int_M\int_M (d(x,y))^p\,d\mu(x,y)&\geqslant \int_A\int_{M\setminus A_h}\bigl(d(x,y)\bigr)^p\,d\mu(x,y)\\&\geqslant h^p\mu\bigl(A\times M\cap M\times(M\setminus A_h)\bigr)\\&\geqslant h^p\bigl(F(A)-G(A_h)\bigr)\geqslant h^{p+1}.\end{aligned}$$

Consequently, $W_p^p(F,G)\geqslant h^{p+1}$, which proves (10.1).

From (10.1) it follows that the topology generated by W_p on $\mathscr{F}_p(M)$ is stronger than the topology of weak convergence. In fact, there is the following characterization (see [30]): $W_p(F_n, F) \to 0$ as $n \to \infty$ if and only if $\pi(F_n, F) \to 0$, and for some (equivalently, all) $x_0 \in M$

$$\int \big(d(x, x_0)\big)^p \, dF_n(x) \to \int \big(d(x, x_0)\big)^p \, dF(x).$$

Thus, convergence in the metric W_p is equivalent to weak convergence on many subspaces of the space $\mathscr{F}_p(M)$.

For $p = 1$ the famous Kantorovich–Rubinstein theorem provides a dual description of the metric W_1 (see [33] and [34]): for all $F, G \in \mathscr{F}_1(M)$

$$W_1(F, G) = \sup \biggl| \int u \, dF - \int u \, dG \biggr|,$$

where the supremum is taken over all functions $u \colon M \to \mathbb{R}$ satisfying the Lipschitz condition $|u(x) - u(y)| \leqslant d(x, y)$, $x, y \in M$.

Therefore, in the case of the real line $M = \mathbb{R}$ with the canonical distance $d(x, y) = |x - y|$, we return to the metric in the space L^1:

$$W_1(F, G) = \zeta_1(F, G) = \|F - G\|_1 = \int_{-\infty}^{\infty} |F(x) - G(x)| \, dx. \tag{10.2}$$

Here F and G on the right-hand side are distribution functions associated with the corresponding probability measures. However, W_p with $p > 1$ does not reduce to the distance in the space L^p, and there is a similar description

$$W_p(F, G) = \biggl(\int_0^1 |F^{-1}(t) - G^{-1}(t)|^p \, dt \biggr)^{1/p}$$

in terms of generalized inverse functions $F^{-1}(t) = \inf\{x \in \mathbb{R} \colon F(x) \geqslant t\}$.

The identity (10.2) suggests that there is possibly a close connection between the Kantorovich and Zolotarev distances. Interesting results in this direction were recently obtained by Rio [35]. In particular, he showed in [35] that for any probability distributions F and G on the real line the estimate

$$W_p(F, G) \leqslant cp \, \zeta_p(F, G)^{1/p}, \qquad p \geqslant 1, \tag{10.3}$$

holds with c an absolute constant.

11. Quadratic Kantorovich distance

In the hierarchy of the metrics W_p with $p > 1$, the particular case $p = 2$ is most popular. The distance W_2 is often regarded as an analogue of the Euclidean distance in the space $\mathscr{F}_2(M)$. Hence, questions connected with estimation of this distance have been the subject of many investigations. The problem of estimating the proximity of F and G in the metric W_2 in terms of the Fourier–Stieltjes transforms is yet to be studied. This observation also applies to other values $p \geqslant 1$, with the exception of the case $p = 1$ and $M = \mathbb{R}$ (in view of the relation (10.2)).

We cite an important result of Talagrand [36] that connects the quadratic Kantorovich distance and the Kullback–Leibler divergence (see also [37]): for all $F \in \mathscr{F}_2(\mathbb{R}^k)$

$$W_2^2(F,\Phi) \leqslant 2\,D(F\|\Phi), \tag{11.1}$$

where Φ is the Gaussian measure on $\mathbb{R}^k$ with density

$$\varphi(x) = \frac{1}{(2\pi)^{k/2}} e^{-\|x\|^2/2}.$$

In the case when the distribution F has density $p(x) = \dfrac{dF(x)}{dx}$ with respect to Lebesgue measure, the estimate (11.1) can be made more concrete:

$$W_2^2(F,\Phi) \leqslant 2 \int_{\mathbb{R}^k} p(x) \log \frac{p(x)}{\varphi(x)}\, dx.$$

The inequality (11.1) is called the transport-entropy inequality. We refer the reader to [38], where other interesting relations are also given for W_2.

Applying (11.1) to smoothed distributions and using the estimate (5.1) for the Kullback–Leibler divergence (Theorem 5.1), one can estimate $W_2(F,\Phi)$ in terms of the characteristic function f of the distribution F. Let us consider the one-dimensional case. We obtain the following theorem, which involves the Fourier–Stieltjes transform

$$g_\alpha(t) = e^{-t^2/2}\biggl(1 + \alpha \frac{(it)^3}{3!}\biggr)$$

of a 'corrected' Gaussian distribution.

Theorem 11.1. *Let F be a probability distribution on the real line with characteristic function f, and let $\int_{-\infty}^{\infty} |x|^3\, dF(x) < \infty$. Then for any $T \geqslant 1$ and $\alpha \in \mathbb{R}$*

$$\begin{aligned} W_2(F,\Phi) \leqslant 4\biggl(\int_{-T}^{T} |f(t) - g_\alpha(t)|^2\, dt\biggr)^{1/4} \\ + 4\biggl(\int_{-T}^{T} |f'''(t) - g_\alpha'''(t)|^2\, dt\biggr)^{1/4} + c\biggl(\frac{1 + Q_T^{1/4}}{T} + |\alpha|\biggr), \end{aligned}$$

where $c > 0$ is an absolute constant and

$$Q_T = \int_{-T}^{T} \bigl(|f''(t)| + |f'(t)| + |f(t)|\bigr)^2 (1 + t^4)\, dt.$$

In applications it is natural to choose T so that the integral Q_T remains bounded, and the quantity α should be chosen to be small of order $1/T$ but not necessarily equal to zero; this can significantly decrease the values of the first two integrals.

In proving the theorem we use the notation $W_2(X,Y)$ instead of $W_2(F,G)$ and $D(X\|Y)$ instead of $D(F\|G)$, where the random variables X and Y have the distributions F and G, respectively.

We put $X_\sigma = X + \sigma Y$ with $\sigma = 1/T$, and we assume that the random variable Y does not depend on X and has a symmetric distribution with characteristic function h such that $h(t) = 0$ for $|t| \geqslant 1$, with $\mathsf{E}|Y|^3 \leqslant C^3$ ($C \geqslant 1$). We can take a normalized convolution power of the triangular characteristic function $h_0(t) = (1 - |t|)_+$, for example

$$h(t) = \frac{h_0^{*6}(6t)}{h_0^{*6}(0)}\,.$$

By convoluting h_0 with itself sufficiently many times, we obtain a positive-definite function that is differentiable the necessary number of times. Consequently, after normalization we are dealing with the characteristic function of a random variable that has finite absolute moments of the necessary order, and we can also control the support of the characteristic function.

By the definition of the quadratic Kantorovich distance,

$$W_2^2(X_\sigma, X) \leqslant \mathsf{E}(X_\sigma - X)^2 = \sigma^2 \mathsf{E} Y^2 \leqslant \frac{C^2}{T^2}\,. \tag{11.2}$$

The triangle inequality for the metric W_2 implies that

$$W_2(X, Z) \leqslant W_2(X_\sigma, Z) + W_2(X_\sigma, X),$$

where it is assumed that the random variable Z has the standard normal distribution Φ. Hence

$$W_2^2(X, Z) \leqslant 2W_2^2(X_\sigma, Z) + 2W_2^2(X_\sigma, X),$$

and from the inequalities (11.1) and (11.2) we derive the smoothing-type inequality

$$W_2^2(X, Z) \leqslant 4\, D(X_\sigma || Z) + \frac{2C^2}{T^2}.$$

To estimate the entropy term in this inequality, we apply (5.1) to the distribution of the random variable X_σ. It has the characteristic function $f_\sigma(t) = f(t)h(\sigma t)$, and thus

$$D(X_\sigma || Z) \leqslant 4\big(\|f_\sigma - g_\alpha\|_2 + \|f_\sigma''' - g_\alpha'''\|_2\big) + \alpha^2,$$

so that we have

$$W_2^2(X, Z) \leqslant 16\left(\|f_\sigma - g_\alpha\|_2 + \|f_\sigma''' - g_\alpha'''\|_2\right) + \frac{2C^2}{T^2} + 4\alpha^2. \tag{11.3}$$

For $|t| \geqslant T$ the equality $f_\sigma(t) = 0$ holds, and we obtain

$$\|f_\sigma - g_\alpha\|_2^2 = \int_{-T}^{T} |f_\sigma(t) - g_\alpha(t)|^2\,dt + \int_{|t| \geqslant T} |g_\alpha(t)|^2\,dt,$$

where the last integral decreases exponentially with respect to T (and even at a higher rate). The same is valid for the derivatives, and hence

$$\int_{|t| \geqslant T} |g_\alpha(t)|^2\,dt \leqslant \frac{C_1(1 + |\alpha|)^2}{T^4} \quad \text{and} \quad \int_{|t| \geqslant T} |g_\alpha'''(t)|^2\,dt \leqslant \frac{C_1(1 + |\alpha|)^2}{T^4}$$

with some constant $C_1 \geqslant 1$. Using the temporary notation

$$\|u\|_T = \left(\int_{-T}^{T} |u(t)|^2\,dt\right)^{1/2},$$

we get from (11.3) that

$$\begin{aligned} W_2^2(X,Z) &\leqslant 16\|f_\sigma - g_\alpha\|_T + 16\|f_\sigma''' - g_\alpha'''\|_T + \frac{C_2(1+|\alpha|)}{T^2} + 4\alpha^2 \\ &\leqslant 16\|f_\sigma - g_\alpha\|_T + 16\|f_\sigma''' - g_\alpha'''\|_T + \frac{C_3}{T^2} + C_4\alpha^2 \end{aligned}$$

with some constants C_j. Moreover, we can approximate f_σ on the interval $[-T,T]$ with the help of f in the sense of the L^2-norm, and by the triangle inequality for the L^2-norm,

$$\begin{aligned} W_2^2(X,Z) \leqslant 16\|f - g_\alpha\|_T + 16\|f''' - g_\alpha'''\|_T + \frac{C_3}{T^2} + C_4\alpha^2 \\ + 16\|f_\sigma - f\|_T + 16\|f_\sigma''' - f'''\|_T. \end{aligned} \tag{11.4}$$

It remains to estimate the last two norms. In view of the symmetry of the distribution of the random variable Y, we have $h'(0) = i\mathsf{E}Y = 0$. Moreover, $|h''(s)| \leqslant \mathsf{E}Y^2 \leqslant C^2$. Consequently, by Taylor's formula $|h(s) - 1| \leqslant C^2 s^2/2$ for all s. It follows that

$$|f_\sigma(t) - f(t)| \leqslant |f(t)|\,|h(\sigma t) - 1| \leqslant \frac{C^2\sigma^2}{2} t^2 |f(t)|$$

and

$$\|f_\sigma - f\|_T \leqslant \frac{C^2\sigma^2}{2}\left(\int_{-T}^{T} t^4 |f(t)|^2\,dt\right)^{1/2} \leqslant \frac{C^2}{2T^2}\sqrt{Q_T}\,. \tag{11.5}$$

Using the inequalities $|h'(s)| \leqslant C|s|$ and $|h^{(r)}(s)| \leqslant \mathsf{E}|Y|^r \leqslant C^r$ $(r = 2,3)$ and differentiating $f_\sigma(t)$ three times, we obtain the similar pointwise estimate

$$\begin{aligned} |f_\sigma'''(t) - f'''(t)| &= |3\sigma f''(t)h'(\sigma t) + 3\sigma^2 f'(t)h''(\sigma t) + \sigma^3 f(t)h'''(\sigma t)| \\ &\leqslant 3C^3\sigma^2\big(|tf''(t)| + |f'(t)| + |f(t)|\big), \end{aligned}$$

which implies that

$$\|f_\sigma''' - f'''\|_T \leqslant 3C^3\sigma^2\left(\int_{-T}^{T}\big(|tf''(t)| + |f'(t)| + |f(t)|\big)^2\,dt\right)^{1/2} \leqslant \frac{3C^3}{T^2}\sqrt{Q_T}\,.$$

Taking this inequality into account along with (11.5), we deduce from (11.4) the required estimate

$$W_2^2(X,Z) \leqslant 16\,\|f - g_\alpha\|_T + 16\|f''' - g_\alpha'''\|_T + \frac{C_3 + 4C^3\sqrt{Q_T}}{T^2} + C_4\alpha^2.$$

12. Smoothing measures with compact support

In the preceding inequalities where smoothing measures were used, the Fourier–Stieltjes transforms of these measures had compact supports. But in some problems it is desirable that the smoothing measures themselves have compact supports. In this case, the corresponding Fourier–Stieltjes transforms can be rapidly decreasing (at infinity), but cannot be concentrated on a finite interval. The problem of the possible rate of decrease was considered by Ingham, who proved the following theorem ([39]; see also [40]). We state it in a slightly different form.

Theorem 12.1. *Let $u\colon [1,\infty) \to [0,\infty)$ be a non-decreasing function such that*

$$I = \int_1^\infty \frac{u(t)}{t}\,dt < \infty.$$

Then for any $c > 1$ there exists a symmetric probability measure H concentrated on the interval $[-cI, cI]$ and with characteristic function f satisfying the inequality

$$|f(t)| \leqslant e^{-tu(t)\log c}, \qquad t \geqslant 4.$$

For example, letting $u(t) = \alpha/(et^\alpha)$ with a parameter α with $0 < \alpha < 1$, and taking $c = e$, we can choose a measure H on $[-1,1]$ with characteristic function f such that

$$|f(t)| \leqslant \exp\biggl\{-\frac{\alpha}{e}t^{1-\alpha}\biggr\}, \qquad t \geqslant 4.$$

Another example $u(t) = \kappa/\log^2(1+t)$ with a suitable value $\kappa > 0$ gives the even more rapid (almost exponential) decrease

$$|f(t)| \leqslant \exp\biggl\{-\frac{\log 2}{2e}\,\frac{t}{\log^2(1+t)}\biggr\}, \qquad t \geqslant 4,$$

and H is again concentrated on $[-1,1]$.

However, it is known that under the assumption of a compact support it is impossible to obtain an inequality of the form $|f(t)| \leqslant Ce^{-ct}$, $t \geqslant t_0$ (with some positive constants c and C). Nevertheless, such an exponential estimate is possible in the integral sense in view of the following elementary theorem.

Theorem 12.2. *For any $T \geqslant 0$ there exists a symmetric probability measure H concentrated on the interval $[-1,1]$ and with characteristic function f satisfying the inequality*

$$\int_{|t|\geqslant T} |f(t)|\,dt \leqslant 2\pi e^2 e^{-T/e}. \tag{12.1}$$

For $T \geqslant 1$

$$\int_{|t|\geqslant T} |f(t)|\,\frac{dt}{|t|} \leqslant 3\,e^{-T/e}. \tag{12.2}$$

As proposed by Ingham, one can take the measure H in Theorem 12.1 to be the distribution of a convergent random series

$$S = c\sum_{n=2}^\infty \frac{u(n)}{n}X_n,$$

where the random variables X_n are independent and uniformly distributed on $(-1,1)$.

To prove Theorem 12.2, it suffices to take the finite sum

$$S_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \qquad n \geqslant 2,$$

with $n = [T/e]$ and $T \geqslant 2e$. In this case $2 \leqslant n \leqslant T/e$, and S_n has the characteristic function $f(t) = \mathsf{E}e^{itS_n} = \left(\frac{\sin(t/n)}{t/n}\right)^n$. Hence,

$$\begin{aligned}\int_T^\infty |f(t)|\,dt &\leqslant \int_T^\infty \left(\frac{n}{t}\right)^n dt = \frac{T}{n-1}e^{-n\log(T/n)} \\ &\leqslant \frac{T}{[T/e]-1}e^{-[T/e]} \leqslant \frac{Te}{[T/e]-1}e^{-T/e} \leqslant 2e^2e^{-T/e}.\end{aligned}$$

In the case $0 \leqslant T \leqslant 2e$, one can take $n=2$ and apply the inverse Fourier transform to $f(t) = \left(\frac{\sin(t/2)}{t/2}\right)^2$. Since S_2 has the triangular density $p(x) = (1-|x|)_+$, we find that $\int_0^\infty f(t)\,dt = \pi p(0) = \pi$. Consequently,

$$\int_0^\infty f(t)\,dt = \pi \leqslant \pi e^2 e^{-T/e}.$$

Combining the two cases, we arrive at (12.1).

Similarly, for $T \geqslant 2e$

$$\int_T^\infty |f(t)|\,\frac{dt}{t} \leqslant \int_T^\infty \left(\frac{n}{t}\right)^n \frac{dt}{t} = \frac{1}{n}e^{-n\log(T/n)} \leqslant \frac{1}{2}e^{-[T/e]} \leqslant \frac{e}{2}^{-T/e}.$$

In the case $1 \leqslant T \leqslant 2e$ one can take $n=1$, which gives

$$\int_T^\infty |f(t)|\,\frac{dt}{t} \leqslant \int_T^\infty \frac{1}{t^2}\,d = \frac{1}{T} \leqslant \frac{3}{2}e^{-T/e}$$

(here we use the fact that $e^{1/e} < 1.5$).

Combining both cases, we obtain (12.2).

The distributions H in Theorem 12.2 were used by Zolotarev to prove an inequality for the Lévy distance (Theorem 3.2).

13. Signed smoothing measures

Let us pass to smoothing measures with additional properties, without keeping the property of positiveness.

Theorem 13.1. *Let $s \geqslant 1$ be an integer. For any $T \geqslant 1$ there exists a symmetric signed measure R on $[-1,1]$ with total variation $\|R\|_{\mathrm{TV}} \leqslant c_s$ such that*

$$R([-1,1]) = 1, \qquad \int_{-1}^{1} x^p\,dR(x) = 0 \quad (p = 1,\dots,s-1) \tag{13.1}$$

and with Fourier–Stieltjes transform f satisfying the inequality

$$\int_{|t|\geqslant T} |f(t)|\,\frac{dt}{|t|} \leqslant 3c_s e^{-T/e}. \tag{13.2}$$

One can set $c_1 = c_2 = 1$, $c_3 = 3$, and $c_s = s\binom{2s}{s-1}$ for $s \geqslant 4$.

In the cases $s = 1$ and $s = 2$ the condition (13.1) is automatically satisfied for the probability measure $R = H$ in Theorem 12.2 (by the symmetry of H; hence $c_1 = c_2 = 1$). For $s = 3$ this condition is satisfied for $p = 1$ but not for $p = 2$. For it to hold it is necessary to drop the property of positiveness of R as a measure, that is, to drop the monotonicity of the associated function $R(x) = R((-\infty, x])$, and to take, for example,

$$R(x) = 2H(x\sqrt{2}\,) - H(x).$$

In this case $\|R\|_{\mathrm{TV}} \leqslant 3$, and thus we can put $c_3 = 3$.

In the general case, we let

$$R(x) = w_1 H\biggl(\frac{x}{b_1}\biggr) + \cdots + w_s H\biggl(\frac{x}{b_s}\biggr),$$

where H is the distribution function in Theorem 12.2 (which we identify with the measure H), and $b_i \neq b_j$ $(i \neq j)$. If all the quantities b_i belong to $(0, 1]$, then R as a measure is concentrated on $[-1, 1]$, and its total variation can be estimated as

$$\|R\|_{\mathrm{TV}} \leqslant \sum_{i=1}^{s} |w_i|. \tag{13.3}$$

The Fourier–Stieltjes transform f of the measure R can be expressed in terms of the characteristic function h of H in the following way: $f(t) = \sum_{i=1}^{s} w_i h(b_i t)$. Hence by Theorem 12.2,

$$\int_{|t|\geqslant T} |f(t)|\,\frac{dt}{|t|} = \sum_{i=1}^{s} |w_i| \int_{|t|\geqslant T/b_i} |h(t)|\,\frac{dt}{|t|} \leqslant 3\sum_{i=1}^{s} |w_i| e^{-T/e}. \tag{13.4}$$

Now we pass to the condition (13.1). It reduces to a linear system in s unknowns $w = (w_1, \dots, w_s)$, which can be written in matrix form as $Vw = e$, where V is the Vandermonde matrix

$$V = \begin{pmatrix} 1 & 1 & \dots & 1 \\ b_1 & b_2 & \dots & b_s \\ \vdots & \vdots & \ddots & \vdots \\ b_1^{s-1} & b_2^{s-1} & \dots & b_s^{s-1} \end{pmatrix}$$

and e is the column $(1, 0, \dots, 0)$. It follows that $w = V^{-1}e$, and we can estimate the right-hand side of (13.3) in terms of the b_i. With this aim in view, we use the following result from [41] on the norm of the inverse of a Vandermonde matrix: if the norm of the $s \times s$ matrix $A = (a_{ij})$ is defined by

$$\|A\| = \max_{1\leqslant i\leqslant s} \sum_{j=1}^{s} |a_{ij}|,$$

then for $A = V^{-1}$ we have

$$\|V^{-1}\| \leqslant \max_{1\leqslant i\leqslant s} \prod_{j\neq i} \frac{1+|b_j|}{|b_i - b_j|}. \tag{13.5}$$

For example, choosing $b_i = i/s$, we obtain the upper estimate

$$\|V^{-1}\| \leqslant \prod_{j=2}^{s} \frac{1+j/s}{j/s-1/s} = \binom{2s}{s-1}.$$

Since $w_i = (Ae)_i = a_{i1}$,

$$\sum_{i=1}^{s} |w_i| = \sum_{i=1}^{s} |a_{i1}| \leqslant \sum_{i=1}^{s}\sum_{j=1}^{s} |a_{ij}| \leqslant s\|A\| \leqslant s\binom{2s}{s-1}.$$

Taking (13.3) and (13.4) into account, we get (13.2) together with the inequality $\|R\|_{\mathrm{TV}} \leqslant s\binom{2s}{s-1}$, which proves the theorem.

14. Analogue of Esseen's inequality for the L^1-metric

In this section we give an example of the use of Theorem 12.2. In particular, we are interested in a variant of Theorem 8.2 in which the L^2-norm of the function $a(t)/t$ on the interval $[-T,T]$ is replaced by the L^1-norm, and the integral containing the derivative is removed.

Let us prove the following theorem. Let $A\colon \mathbb{R}\to\mathbb{R}$ be a function of bounded variation with Fourier–Stieltjes transform

$$a(t) = \int_{-\infty}^{\infty} e^{itx}\,dA(x), \qquad t\in\mathbb{R},$$

and let $A(-\infty)=0$.

Theorem 14.1. *If* $1 \leqslant \beta-\alpha \leqslant T\log T$ $(T\geqslant 1)$, *then*

$$\int_{\alpha}^{\beta} |A(x)|\,dx \leqslant \frac{\beta-\alpha}{2\pi}\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt + c\|A\|_{\mathrm{TV}}\frac{\log T}{T}\,, \tag{14.1}$$

where c *is an absolute positive constant.*

In a somewhat weaker form when $A = F - G$ is the difference of two distribution functions, such an estimate can be obtained with the help of the Zolotarev inequality for the Lévy distance (Theorem 3.2). Indeed, if $h > L(F,G)$, then for all x

$$\begin{aligned} F(x)-G(x) &\leqslant (G(x+h)-G(x)) + h,\\ G(x)-F(x) &\leqslant (F(x+h)-F(x)) + h, \end{aligned}$$

and it follows that

$$|F(x)-G(x)| \leqslant (F(x+h)-F(x)) + (G(x+h)-G(x)) + h.$$

Integrating this inequality over the interval $[\alpha, \beta]$ and then letting h tend to $L(F, G)$, we find that

$$\int_\alpha^\beta |F(x) - G(x)|\, dx \leqslant (1 + (\beta - \alpha))L(F, G).$$

Thus, for all $T > 1.3$ the inequality (3.3) implies that

$$\int_\alpha^\beta |F(x) - G(x)|\, dx \leqslant \frac{1 + (\beta - \alpha)}{2\pi} \int_{-T}^{T} \left| \frac{f(t) - g(t)}{t} \right| dt + 2e\bigl(1 + (\beta - \alpha)\bigr) \frac{\log T}{T}\,,$$

where f and g are the characteristic functions of the distributions F and G. Hence, an advantage of the estimate (14.1) is the absence of the coefficient $\beta - \alpha$ on the right-hand side.

To prove Theorem 14.1 we need an auxiliary inequality, which is of independent interest.

Lemma 14.2. *Let $A\colon [\alpha, \beta] \to \mathbb{R}$ be a function of bounded variation. Then for any integer $N \geqslant 1$*

$$\int_\alpha^\beta |A(x)|\, dx \leqslant \sum_{k=1}^{N} \left| \int_{x_{k-1}}^{x_k} A(x)\, dx \right| + \frac{\beta - \alpha}{N} \|A\|_{\mathrm{TV}}, \tag{14.2}$$

where $x_k = \alpha + (\beta - \alpha)k/N$.

This inequality was proved in [42] for the difference of distribution functions (with $\|A\|_{\mathrm{TV}}$ replaced by 2). The general case is similar; let I denote the collection of indices $k = 1, \dots, N$ such that the function $A(x)$ does not change sign in the kth interval $\Delta_k = (x_{k-1}, x_k)$. The other indices form the complementary subset $J \subset \{1, \dots, N\}$. Then for all $k \in I$

$$\int_{\Delta_k} |A(x)|\, dx = \left| \int_{\Delta_k} A(x)\, dx \right|.$$

But if $k \in J$, then obviously

$$\sup_{x \in \Delta_k} |A(x)| \leqslant \sup_{x, y \in \Delta_k} |A(x) - A(y)| \leqslant |A|(\Delta_k),$$

where $|A|$ is the variation of the function A, regarded as a positive measure on $[\alpha, \beta]$. In this case we obtain

$$\int_{\Delta_k} |A(x)|\, dx \leqslant |A|(\Delta_k)|\Delta_k|, \qquad |\Delta_k| = \frac{\beta - \alpha}{N}\,.$$

Combining both estimates, we conclude that the integral $\int_\alpha^\beta |A(x)|\, dx$ does not exceed the quantity

$$\sum_{k \in I} \left| \int_{\Delta_k} A(x)\, dx \right| + \sum_{k \in J} |A|(\Delta_k)|\Delta_k| \leqslant \sum_{k=1}^{N} \left| \int_{\Delta_k} A(x)\, dx \right| + \frac{\beta - \alpha}{N} \sum_{k=1}^{N} |A|(\Delta_k).$$

In view of the fact that the measure $|A|$ is additive, the last sum does not exceed $\|A\|_{\mathrm{TV}}$. The proof of the lemma is complete. $\square$

Passing to the proof of (14.1), we need to estimate the integral on the right-hand side of (14.2). It can be assumed that $\displaystyle\int_{-T}^{T}\left|\frac{a(t)}{t}\right| dt<\infty$. We consider a function of the form

$$U(x)=\frac{1}{2h}\int_{x-h}^{x+h} A(y)\,dy, \qquad x\in\mathbb{R}, \quad h>0,$$

which is the convolution of the measure A with the uniform distribution on an interval $(-h,h)$ (the parameter h will be chosen later). It has the Fourier–Stieltjes transform $\dfrac{\sin(th)}{th}a(t)$.

Fixing another parameter $\sigma>0$, which will be chosen depending on T, we consider the convolution

$$U_\sigma(x)=\int_{-\infty}^{\infty} U(x-\sigma y)\,dH(y),$$

where H is the probability measure in Theorem 12.2. Since this measure is concentrated on $[-1,1]$, we immediately obtain

$$|U_\sigma(x)-U(x)|\leqslant \sup_{|z|\leqslant\sigma}|U(x-z)-U(x)|\leqslant\frac{\sigma}{2h}|A|(x-h-\sigma,x+h+\sigma), \quad (14.3)$$

where the last factor on the right-hand side denotes the variation of A on $(x-h-\sigma, x+h+\sigma)$.

On the other hand, U_σ has the Fourier–Stieltjes transform

$$g(t)=f(\sigma t)\frac{\sin(th)}{th}a(t),$$

where f is the characteristic function of the measure H, and $U_\sigma(-\infty)=0$. The function g is integrable, and hence the inverse Fourier–Stieltjes transform can be applied: for all x

$$U_\sigma(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{g(t)}{-it}\,dt.$$

Using the inequality $\sup_t|a(t)|\leqslant\|A\|_{\mathrm{TV}}$ outside the interval $[-T,T]$, we obtain

$$|U_\sigma(x)|\leqslant\frac{1}{2\pi}\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt+\frac{1}{2\pi}\|A\|_{\mathrm{TV}}\int_{|t|\geqslant\sigma T}\left|\frac{f(t)}{t}\right|dt.$$

Here we put $\sigma=(2e\log T)/T$ and use (12.2) with $2e\log T$ instead of T to estimate the last integral. Then we arrive at the estimate

$$|U_\sigma(x)|\leqslant\frac{1}{2\pi}\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt+\frac{3}{2\pi}\,\frac{\|A\|_{\mathrm{TV}}}{T^2}$$

and from (14.3) we get that

$$|U(x)|\leqslant\frac{1}{2\pi}\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt+\frac{3}{2\pi}\,\frac{\|A\|_{\mathrm{TV}}}{T^2}+\frac{\sigma}{2h}|A|(x-\sigma-h,x+\sigma+h).$$

Therefore, by the definition of the function U and the choice of σ, we find that for all $x \in \mathbb{R}$, $h > 0$, and $T \geqslant 1$

$$\left|\int_{x-h}^{x+h} A(y)\,dy\right| \leqslant \frac{h}{\pi}\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt + \frac{3}{\pi}\,\frac{h}{T^2}\|A\|_{\mathrm{TV}} + \frac{2e\log T}{T}\varepsilon\biggl(x, h + \frac{2e\log T}{T}\biggr), \tag{14.4}$$

where

$$\varepsilon(x,r) = |A|(x-r, x+r).$$

We now return to Lemma 14.2 and use the same partition of $[\alpha,\beta]$ into intervals $\Delta_k = (x_{k-1}, x_k)$ with endpoints $x_k = \alpha + (\beta-\alpha)k/N$. If we apply (14.4) to the points $z_k = (x_{k-1}+x_k)/2$ and $h = (\beta-\alpha)/(2N)$ and sum over all $k = 1,\dots,N$, then (14.2) implies that

$$\begin{aligned}\int_{\alpha}^{\beta}|A(x)|\,dx \leqslant \frac{\beta-\alpha}{2\pi}\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt + \frac{3}{2\pi}\,\frac{\beta-\alpha}{T^2}\|A\|_{\mathrm{TV}} \\ + \frac{\beta-\alpha}{N}\|A\|_{\mathrm{TV}} + \frac{2e\log T}{T}\sum_{k=1}^{N}\varepsilon\biggl(z_k, h + \frac{2e\log T}{T}\biggr).\end{aligned} \tag{14.5}$$

Note that $\sum_{k=1}^{N}\varepsilon(z_k, lh) \leqslant l\|A\|_{\mathrm{TV}}$ for any integer $l \geqslant 1$. Thus, the sum in (14.5) does not exceed the quantity

$$\biggl(\frac{2e\log T}{Th} + 2\biggr)\|A\|_{\mathrm{TV}}.$$

Consequently, taking the inequality $\beta-\alpha \leqslant T\log T$ into account, we arrive at the estimate

$$\begin{aligned}\int_{\alpha}^{\beta}|A(x)|\,dx \leqslant \frac{\beta-\alpha}{2\pi}\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt \\ + \|A\|_{\mathrm{TV}}\biggl[\frac{3}{2\pi}\,\frac{\log T}{T} + \frac{\beta-\alpha}{N} + \frac{2e\log T}{T}\biggl(\frac{2e\log T}{Th} + 2\biggr)\biggr].\end{aligned} \tag{14.6}$$

Letting $N = [(\beta-\alpha)T/\log T] + 1$, we obtain $(\beta-\alpha)/N \leqslant (\log T)/T$, and if $\beta-\alpha \geqslant 1$, then

$$Th = T\frac{\beta-\alpha}{2N} \geqslant T\frac{\beta-\alpha}{2(\beta-\alpha)T/\log T + 2} \geqslant \frac{1}{3}\log T.$$

Thus, up to the factor $(\log T)/T$ the expression in square brackets on the right-hand side of (14.6) does not exceed $3/(2\pi) + 1 + 2e(6e+2) < 102$, and we obtain (14.1) with $c = 102$. Theorem 14.1 is proved. $\square$

15. Variants of the Berry–Esseen inequality

Again let $A\colon \mathbb{R} \to \mathbb{R}$ be a function of bounded variation, with $A(-\infty) = 0$. We return to estimations of the L^∞-norm

$$\|A\| = \sup_x |A(x)|$$

in terms of the Fourier–Stieltjes transform

$$a(t) = \int_{-\infty}^{\infty} e^{itx}\,dA(x).$$

It we let $F = 0$ and $G = A$, then, under the assumption that $|A'(x)| \leqslant L_1$ for all x, the Berry–Esseen inequality (Theorem 2.1) gives the estimate

$$\|A\| \leqslant c\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt + c'\,\frac{L_1}{T}\,, \qquad T > 0, \tag{15.1}$$

with absolute constants $c, c' > 0$. Interestingly, if we strengthen the property of smoothness in terms of higher-order derivatives of the function A, then (15.1) can be improved significantly. In particular, the following statement holds.

Theorem 15.1. *If a function A is twice differentiable and $\sup_x |A''(x)| \leqslant L_2$, then for all $T > 0$*

$$\|A\| \leqslant c\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt + c'\,\frac{L_2}{T^2}\,, \tag{15.2}$$

where c and c' are absolute positive constants.

It is also of interest to find out whether this estimate can be made local, when smoothness properties of the function A are known only in a neighbourhood of a given point x. It turns out that a similar statement can be obtained in this case.

Theorem 15.2. *Assume that a function A is differentiable s times in a neighbourhood $\Delta\colon |z - x| < se(\log T)/T$ of a given point x (where $T > 1$ and $s \geqslant 1$ is an integer) and that*

$$\sup_{z\in\Delta} |A^{(s)}(z)| \leqslant L_s(x, T). \tag{15.3}$$

Then

$$|A(x)| \leqslant \frac{1}{2\pi}\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt + c_s\|A\|_{\mathrm{TV}}\frac{1}{T^s} + c_s L_s(x, T)\left(\frac{\log T}{T}\right)^s, \tag{15.4}$$

where c_s is a constant depending only on s.

We stress that such estimates are hardly of value in problems concerning, for example, the rate of convergence in the central limit theorem when one is considering functions $A = F - G$ with smooth G but generally discontinuous F. Nevertheless, 'discrete' analogues of the inequalities (15.2) and (15.4) can be derived for such purposes, and the condition (15.3) should be stated in terms of difference operators of order s. Following this line of investigation, one can study, for example, Edgeworth expansions for binomial distributions. For lack of space we do not touch upon such generalizations here.

We start with a proof of Theorem 15.2, assuming that $\displaystyle\int_{-T}^{T}\left|\frac{a(t)}{t}\right|dt < \infty$. The inequality (15.4) is based on the smoothing of A by the signed measure R in Theorem 13.1: we consider the convolution

$$A_\sigma(x) = \int_{-\infty}^{\infty} A(x - \sigma y)\,dR(y) \tag{15.5}$$

with the parameter $\sigma = (se \log T)/T$. Since R is concentrated on the interval $[-1, 1]$ and $R([-1, 1]) = 1$, we have

$$A_\sigma(x) - A(x) = \int_{-1}^{1} (A(x - \sigma y) - A(x))\, dR(y).$$

Expanding $A(x - \sigma y)$ by Taylor's formula in powers of y and using (13.1), we get that

$$|A_\sigma(x) - A(x)| \leqslant \frac{\sigma^s}{s!} \sup_{z \in \Delta} |A^{(s)}(z)| \leqslant \frac{\sigma^s}{s!} L_s(x, T). \tag{15.6}$$

On the other hand, A_σ has the Fourier–Stieltjes transform $g(t) = f(\sigma t) a(t)$, where f is the Fourier–Stieltjes transform of the measure R, and $A_\sigma(-\infty) = 0$. Hence we can apply the inverse Fourier–Stieltjes transform:

$$A_\sigma(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{g(t)}{-it}\, dt. \tag{15.7}$$

Using the inequality $\sup_t |a(t)| \leqslant \|A\|_{\mathrm{TV}}$ outside $[-T, T]$ and then applying (13.2), we find that

$$\begin{aligned} |A_\sigma(x)| &\leqslant \frac{1}{2\pi} \int_{-T}^{T} \left| \frac{a(t)}{t} \right| dt + \frac{\|A\|_{\mathrm{TV}}}{2\pi} \int_{|t| \geqslant \sigma T} \left| \frac{f(t)}{t} \right| dt \\ &\leqslant \frac{1}{2\pi} \int_{-T}^{T} \left| \frac{a(t)}{t} \right| dt + \frac{\|A\|_{\mathrm{TV}}}{2\pi} \frac{c_s}{T^s} \end{aligned}$$

with the constant c_s from Theorem 13.1.

It remains to combine this inequality with (15.6):

$$|A(x)| \leqslant \frac{1}{2\pi} \int_{-T}^{T} \left| \frac{a(t)}{t} \right| dt + \frac{\|A\|_{\mathrm{TV}}}{2\pi} \frac{c_s}{T^s} + \frac{(se)^s}{s!} L_s(x, T) \left(\frac{\log T}{T} \right)^s.$$

Since $c_1 = c_2 = 1$, we get, in particular, that

$$|A(x)| \leqslant \frac{1}{2\pi} \int_{-T}^{T} \left| \frac{a(t)}{t} \right| dt + \frac{\|A\|_{\mathrm{TV}}}{2\pi} \frac{1}{T} + e L_1(x, T) \frac{\log T}{T},$$

$$|A(x)| \leqslant \frac{1}{2\pi} \int_{-T}^{T} \left| \frac{a(t)}{t} \right| dt + \frac{\|A\|_{\mathrm{TV}}}{2\pi} \frac{1}{T^2} + 2e^2 L_2(x, T) \left(\frac{\log T}{T} \right)^2.$$

In the general case $c_s \leqslant s\binom{2s}{s-1}$, and we arrive at the required inequality (15.4) with $c_s = c^s$, where c is an absolute constant.

It is somewhat simpler to prove (15.2); we can follow the standard reasoning used in deriving the Berry–Esseen inequality. As a smoothing measure in (15.5), we take the probability measure R with the triangular characteristic function $f(t) = (1 - |t|)_+$, that is, with the density

$$\psi(x) = \frac{1}{2\pi} \left(\frac{\sin(x/2)}{x/2} \right)^2,$$

and we let $\sigma = 1/T$. Then we again have (15.7), and it follows immediately that

$$\|A_\sigma\| \leqslant \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\frac{a(t)}{t}\right| f(\sigma t)\, dt \leqslant \frac{1}{2\pi} \int_{-T}^{T} \left|\frac{a(t)}{t}\right| dt. \tag{15.8}$$

On the other hand, starting from (15.5), for fixed $l > 0$ we write

$$A_\sigma(x) - A(x) = \int_{-l}^{l} (A(x-\sigma y) - A(x))\, dR(y) + \int_{|y|>l} (A(x-\sigma y) - A(x))\, dR(y).$$

Expanding $A(x - \sigma y)$ up to the quadratic term by Taylor's formula and using the symmetry of the measure R, we easily see that the first integral can be estimated as follows:

$$\left|\int_{-l}^{l} (A(x-\sigma y) - A(x))\, dR(y)\right| \leqslant \int_{-l}^{l} \frac{(\sigma y)^2}{2} L_2\, dR(y) = \frac{l^3}{3}\, \frac{L_2}{T^2}\,.$$

We estimate the modulus of the second integral simply by $2\gamma\|A\|$, where $\gamma = 1 - R[-l, l]$. Using (15.8), we obtain the inequality

$$\|A\| \leqslant \frac{1}{2\pi} \int_{-T}^{T} \left|\frac{a(t)}{t}\right| dt + \frac{l^3}{3}\, \frac{L_2}{T^2} + 2\gamma\|A\|,$$

which in the case $\gamma < 1/2$ gives the required estimate

$$\|A\| \leqslant \frac{1}{2\pi(1-2\gamma)} \int_{-T}^{T} \left|\frac{a(t)}{t}\right| dt + \frac{l^3}{3(1-2\gamma)}\, \frac{L_2}{T^2}\,.$$

To fix the numerical values of the constants, we put $l = 3\pi/2$, for example. In this case

$$\begin{aligned} 2\gamma &= 4\int_{3\pi/2}^{\infty} \psi(x)\, dx = \frac{4}{\pi} \int_{3\pi/2}^{\infty} \frac{1-\cos x}{x^2}\, dx \\ &\leqslant \frac{4}{\pi} \int_{3\pi/2}^{5\pi/2} \frac{1}{x^2}\, dx + \frac{4}{\pi} \int_{5\pi/2}^{\infty} \frac{2}{x^2}\, dx = \frac{64}{15\pi^2} = 0.4323\ldots\,. \end{aligned}$$

Consequently, the inequality (15.2) holds with the constants

$$c = \frac{1}{2\pi(1-2\gamma)} = 0.2804\ldots \quad \text{and} \quad c' = \frac{l^3}{3(1-2\gamma)} = \frac{9\pi^3}{8(1-2\gamma)} = 61.44\ldots\,.$$

16. Smoothing with a polynomial weight

We now return to functions of bounded variation of the form $A = F - G$ with a non-decreasing function F and with smoothness-type conditions on G. We will be interested in a generalization of Theorem 2.1 for the Kolmogorov distance with a polynomial weight, namely, for

$$\rho_s(F, G) = \sup_x (1 + |x|^s)|F(x) - G(x)|,$$

where $s \geqslant 0$ is a given integer (so that $\rho_0 = \rho$).

The need to study such distances comes from the importance of non-uniform estimates

$$|F(x) - G(x)| \leqslant \frac{c}{1 + |x|^s}, \qquad x \in \mathbb{R}, \tag{16.1}$$

where we can put $c = \rho_s(F, G)$. For example, the particular case $s = 2$ lets us estimate the distance in the L^1-metric: integrating (16.1) over the whole real line, we obtain the relation

$$W_1(F, G) \leqslant \pi \rho_2(F, G).$$

It also remains valid for the distance in the L^p-metric for all $p \geqslant 1$. Note that for $p > 1$ we can use (16.1) also with $s = 1$, and then we get that

$$\|F - G\|_p \leqslant \left(\frac{2}{p-1}\right)^{1/p} \rho_1(F, G).$$

Before estimating ρ_s in terms of the corresponding Fourier–Stieltjes transforms of the functions F and G, we fix in this section a general relation of the type (2.2) between the distance $\rho_s(F, G)$ and the L^∞-norm of the smoothed function

$$A_s(x) = x^s (F(x) - G(x)).$$

For this purpose, it is natural to require that F and G have finite absolute moments of order s, that is,

$$\int_{-\infty}^{\infty} |x|^s \, dF(x) < \infty, \qquad \int_{-\infty}^{\infty} |x|^s \, |dG(x)| < \infty,$$

where $|G|$ denotes the variation of G as a positive measure on the real line. In this case, A_s is a function of bounded variation, and $A_s(-\infty) = A_s(\infty) = 0$.

Let us fix a distribution function H, put $H_T(x) = H(Tx)$ for $T > 0$, and choose an $l > 0$ such that the condition $1 - H([-l, l]) \leqslant 1/4$ is satisfied.

Lemma 16.1. *Assume that a distribution function F and a function G of bounded variation have finite absolute moments of integer order $s \geqslant 1$ and that $G(-\infty) = 0$ and $G(\infty) = 1$. If G is differentiable and its derivative satisfies the inequality*

$$\sup_x \, (1 + |x|^s) \, |G'(x)| \leqslant L, \tag{16.2}$$

then for any $T \geqslant 1$

$$\|A_0\| + \|A_s\| \leqslant c\left(\|A_0 * H_T\| + \|A_s * H_T\| + \frac{L}{T}\right), \tag{16.3}$$

where the constant $c = c(s, l)$ depends only on s and l.

Proof. We prove the lemma by standard but routine arguments (see [3], Chap. VI, Lemma 8). First we consider the non-polynomial case $s = 0$ and derive the inequality (2.2). Letting $\sigma = 1/T$, we consider the convolution

$$I(x) \equiv (A * H_T)(x) = \int_{-\infty}^{\infty} A(x - \sigma y) \, dH(y) = I_0(x) + I_1(x),$$

where $A = A_0 = F - G$ and

$$I_0(x) = \int_{|y|\leqslant l} A(x-\sigma y)\,dH(y), \quad I_1(x) = \int_{|y|>l} A(x-\sigma y)\,dH(y).$$

We have $|I_1(x)| \leqslant \gamma\|A\|$ with the coefficient $\gamma = 1 - H([-l,l])$. For an estimate of the first integral, the monotonicity of F and the Lipschitz property of G let us use the inequalities

$$A(x-\sigma y) \geqslant A(x-\sigma l) - 2\sigma lL \quad \text{and} \quad -A(x-\sigma y) \geqslant -A(x+\sigma l) - 2\sigma lL,$$

which imply the estimate

$$|I_0(x)| \geqslant (1-\gamma)\max\{A(x-\sigma l), -A(x+\sigma l)\} - 2(1-\gamma)\sigma lL.$$

Since $|I(x)| \geqslant |I_0(x)| - |I_1(x)|$, we obtain

$$\|I\| \geqslant (1-\gamma)\max\{A(x-\sigma l), -A(x+\sigma l)\} - 2(1-\gamma)\sigma lL - \gamma\|A\|.$$

Taking the supremum over all x, we arrive at the estimate $\|I\| \geqslant (1-2\gamma)\|A\| - 2(1-\gamma)\sigma lL$, that is,

$$\|A\| \leqslant \frac{1}{1-2\gamma}\|A * H_T\| + 2l\frac{1-\gamma}{1-2\gamma}\,\frac{L}{T},$$

which coincides exactly with (2.2). In particular, if $\gamma \leqslant 1/4$, then this estimate yields

$$\|A\| \leqslant 2\|A * H_T\| + 3l\frac{L}{T}. \tag{16.4}$$

Now let $s \geqslant 1$. Fixing an arbitrary value $\varepsilon \in (0,1/2]$, we choose a point x_0 such that

$$|A_s(x_0)| \geqslant (1-\varepsilon)\|A_s\| \geqslant \frac{1}{2}\|A_s\|. \tag{16.5}$$

Without loss of generality we can make the two assumptions

$$|x_0| \geqslant 4sl \quad \text{and} \quad |A_s(x_0)| \geqslant 2^{s+3}\sigma lL. \tag{16.6}$$

Indeed, if the first condition fails, then by (16.4) and (16.5) we have

$$\begin{aligned}\|A_s\| &\leqslant 2|A_s(x_0)| \leqslant 2(4sl)^s|A(x_0)| \\ &\leqslant 2(4sl)^s\|A\| \leqslant 4(4sl)^s\|A * H_T\| + 6l(4sl)^s\frac{L}{T},\end{aligned}$$

which automatically implies (16.3), namely,

$$\|A\| + \|A_s\| \leqslant \bigl(2 + 4(4sl)^s\bigr)\|A * H_T\| + \bigl(3l + 6l(4sl)^s\bigr)\frac{L}{T}. \tag{16.7}$$

If the second condition fails, then $\|A_s\| \leqslant 2|A_s(x_0)| \leqslant 2^{s+4}lL/T$, which gives us the similar estimate

$$\|A\| + \|A_s\| \leqslant 2\|A * H_T\| + (3 + 2^{s+4})l\frac{L}{T}. \tag{16.8}$$

We consider a convolution analogous to the one at the beginning of the proof:

$$I(x) \equiv (A_s * H_T)(x) = \int_{-\infty}^{\infty} A_s(x-\sigma y)\,dH(y) = I_0(x) + I_1(x),$$

where

$$I_0(x) = \int_{|y|\leqslant l} A_s(x-\sigma y)\,dH(y), \qquad I_1(x) = \int_{|y|>l} A_s(x-\sigma y)\,dH(y).$$

Again, $|I_1(x)| \leqslant \gamma\|A_s\|$, with $\gamma = 1 - H([-l,l])$, so for all x

$$\|I\| \geqslant |I(x)| \geqslant |I_0(x)| - \gamma\,\|A_s\|. \tag{16.9}$$

Next we estimate $|I_0(x)|$ from below either at the point $x = x_0 - \sigma l$ or at $x = x_0 + \sigma l$, depending on the sign of $A(x_0)$. For definiteness we assume that $A(x_0) < 0$.

For all $z \in [x_0 - 2\sigma l, x_0]$ there exists a point z_0 in the same interval such that $G(z) = G(x_0) + (z - x_0)G'(z_0)$. We have $|z_0| \geqslant |x_0|$ if $x_0 < 0$, and $z_0 \geqslant x_0/2$ if $x_0 > 0$. Indeed, in the latter case the worst variant is attained for $z_0 = x_0 - 2\sigma l$. But $x_0 \geqslant 4\sigma l$ and $\sigma \leqslant 1$, which implies that $z_0 \geqslant x_0/2$. Therefore, by the condition (16.2),

$$|G(z) - G(x_0)| = |z - x_0|\,|G'(z_0)| \leqslant 2\sigma l \frac{L}{|z_0|^s} \leqslant 2^{s+1}\sigma l \frac{L}{|x_0|^s}\,.$$

Using the monotonicity of F, we get that

$$\begin{aligned} A(z) = F(z) - G(z) &\leqslant F(x_0) - G(x_0) + 2^{s+1}\sigma l \frac{L}{|x_0|^s} \\ &= A(x_0) + 2^{s+1}\sigma l \frac{L}{|x_0|^s} < \frac{3}{4}A(x_0), \end{aligned}$$

where we have used the second condition in (16.6) in the last step. Furthermore, the inequality $|z|^s \geqslant |x_0|^s$ holds in the case $x_0 < 0$, while in the case $x_0 > 0$ the first condition in (16.6) and the assumption $\sigma \leqslant 1$ give us that

$$z^s \geqslant (x_0 - 2\sigma l)^s \geqslant (x_0 - 2l)^s \geqslant \left(1 - \frac{1}{2s}\right)^s x_0^s \geqslant \frac{1}{\sqrt{e}}\, x_0^s.$$

In both cases $|z|^s \geqslant |x_0|^s/\sqrt{e}$, and hence $A_s(z) = z^s A(z)$ does not change sign in the interval $x_0 - 2\sigma l \leqslant z \leqslant x_0$ and satisfies there the inequality

$$|A_s(z)| \geqslant \frac{3}{4\sqrt{e}}|A_s(x_0)|.$$

It follows that

$$|I_0(x)| = \int_{|y|\leqslant l} |A_s(x-\sigma y)|\,dH(y) \geqslant (1-\gamma)\frac{3}{4\sqrt{e}}|A_s(x_0)| \geqslant \frac{1-\gamma}{1-\varepsilon}\,\frac{3}{4\sqrt{e}}\|A_s\|,$$

where we used the estimate (16.5) in the last step.

Arguing similarly, we can show that these inequalities hold also in the case $A(x_0) > 0$ for the point $x = x_0 + \sigma l$ and the interval $x_0 \leqslant z \leqslant x_0 + 2\sigma l$. Therefore, letting ε tend to zero, we arrive in both cases at the same estimate

$$|I_0(x)| \geqslant \frac{3}{4\sqrt{e}}(1-\gamma)\|A_s\|.$$

Using it in (16.9), we obtain

$$\|I\| \geqslant \left(\frac{3}{4\sqrt{e}}(1-\gamma) - \gamma\right)\|A_s\| \geqslant 0.091\|A_s\|,$$

where in the last inequality we assume that $\gamma \leqslant 1/4$. Consequently, in this case $\|A_s\| \leqslant 11\|I\| = 11\|A_s * H_T\|$, and in view of (16.4) we have

$$\|A\| + \|A_s\| \leqslant 2\|A * H_T\| + 11\|A_s * H_T\| + 3l\frac{L}{T}\,.$$

By (16.7) and (16.8), we obtain (16.3), and Lemma 16.1 is proved. □

17. General non-uniform estimates

The right-hand side of the inequality (16.3) in Lemma 16.1 can be estimated further in terms of the Fourier–Stieltjes transforms of the functions F and G. In this step we can consider broader classes of functions.

Let A be a function of bounded variation with finite absolute moment of integer order $s \geqslant 1$ (for the variation $|A|$, regarded as a measure on the line), and with $A(-\infty) = A(\infty) = 0$. In this case, the corresponding Fourier–Stieltjes transform

$$a(t) = \int_{-\infty}^{\infty} e^{itx}\, dA(x) = -it\int_{-\infty}^{\infty} e^{itx}\, A(x)\, dx$$

has s continuous derivatives, and the function

$$A_s(x) = x^s A(x)$$

also has bounded variation, with $A_s(-\infty) = A_s(\infty) = 0$.

Theorem 17.1. *For each distribution function H with characteristic function h,*

$$\|A_s * H\| \leqslant \frac{2}{\pi} \sup_{x\in\mathbb{R},\ \kappa\geqslant 1}\left|\int_{-\infty}^{\infty} e^{-itx}\frac{a^{(s)}(t)}{t}h(\kappa t)\, dt\right|. \tag{17.1}$$

In particular, in the absence of any smoothing (or when H is a unit weight at zero, with $h(t) = 1$), we have

$$\|A_s\| \leqslant \frac{2}{\pi} \sup_x\left|\int_{-\infty}^{\infty} e^{-itx}\frac{a^{(s)}(t)}{t}\, dt\right|. \tag{17.2}$$

However, as in the ordinary Berry–Esseen inequality, smoothing reduces the problem of estimating $\|A_s\|$ to the problem of estimating $a^{(s)}(t)$ on an interval $|t| \leqslant T$ with an error of order $1/T$.

The integrals in (17.1) and (17.2) are always finite, but they are understood in the sense of their principal values, as limits of the integrals over the set $\varepsilon < |t| < T$ as $\varepsilon \downarrow 0$ and $T \uparrow \infty$. Here we can use the following variant of inversion for Fourier–Stieltjes transforms: if B is a function of bounded variation with the Fourier–Stieltjes transform b, then for all x

$$\frac{1}{\pi} \lim_{\varepsilon \downarrow 0,\ T \uparrow \infty} \int_{\varepsilon<|t|<T} \frac{e^{-itx}}{-it} b(t)\, dt = \int_{-\infty}^{\infty} \left(1_{\{y<x\}} - 1_{\{y>x\}}\right) dB(y). \tag{17.3}$$

If B is right-continuous and $B(-\infty) = 0$ (which is always assumed), then the integrals on the right-hand side of (17.3) are connected with the L^∞-norm of the function B as follows:

$$\|B\| \leqslant \sup_x \left| \int_{-\infty}^{\infty} \left(1_{\{y<x\}} - 1_{\{y>x\}}\right) dB(y) \right| \leqslant 3\|B\|. \tag{17.4}$$

Indeed, we denote these integrals by $I(x)$ and let $M = \sup_x |I(x)|$. Letting x tend to infinity, we arrive at the estimate $M \geqslant |B(\infty)|$. Since $2B(x) = B(\infty) + I(x)$ at each point x at which B is continuous, we immediately obtain $|B(x)| \leqslant M$, that is, the left-hand inequality in (17.4). Moreover, $|I(x)| \leqslant 2|B(x)| + |B(\infty)| \leqslant 3\|B\|$, which yields the right-hand inequality.

To prove the theorem we need two elementary equalities.

Lemma 17.2. *For $s \geqslant 1$ the function A_s has the Fourier–Stieltjes transform*

$$a_s(t) = i^{-s} t \left(\frac{a(t)}{t}\right)^{(s)} = i^{-s} \int_0^1 \left(a^{(s)}(t) - a^{(s)}(\eta t)\right) d\eta^s, \qquad t \neq 0. \tag{17.5}$$

Proof. The first relation (which is also true for $s = 0$) can be obtained if we differentiate the equality $\dfrac{a(t)}{it} = -\displaystyle\int_{-\infty}^{\infty} e^{itx} A(x)\, dx$ s times with respect to t and then integrate by parts:

$$\left(\frac{a(t)}{it}\right)^{(s)} = -i^s \int_{-\infty}^{\infty} e^{itx} A_s(x)\, dx = \frac{i^s}{it} \int_{-\infty}^{\infty} e^{itx}\, dA_s(x) = \frac{i^s}{it} a_s(t).$$

However, this argument is not quite rigorous, because A_s is not necessarily Lebesgue integrable. But the function A_{s-1} is integrable, therefore

$$\left(\frac{a(t)}{it}\right)^{(s-1)} = -i^{s-1} \int_{-\infty}^{\infty} e^{itx} A_{s-1}(x)\, dx, \qquad t \neq 0. \tag{17.6}$$

Since $dA_s(x) = x^s\, dA(x) + sA_{s-1}(x)\, dx$, we have another identity:

$$\int_{-\infty}^{\infty} e^{itx}\, dA_s(x) = \int_{-\infty}^{\infty} e^{itx} x^s\, dA(x) + s \int_{-\infty}^{\infty} e^{itx} A_{s-1}(x)\, dx,$$

that is, by (17.6),

$$a_s(t) = i^{-s} \left[a^{(s)}(t) - s\left(\frac{a(t)}{t}\right)^{(s-1)}\right].$$

Thus, it remains to establish the equality $a^{(s)}(t) - s(a(t)/t)^{(s-1)} = t(a(t)/t)^{(s)}$, or, what is the same in terms of $v(t) = a(t)/t$, the equality

$$(tv(t))^{(s)} = tv^{(s)}(t) + sv^{(s-1)}(t),$$

which is obvious.

The second relation in (17.5) follows from the identity

$$\left(\frac{a(t)}{t}\right)^{(s)} = \frac{1}{t}\int_0^1 \bigl(a^{(s)}(t) - a^{(s)}(\eta t)\bigr)\,d\eta^s, \qquad t \neq 0, \tag{17.7}$$

which is valid for any function a that is continuously differentiable s times and satisfies $a(0) = 0$. Without loss of generality we can assume here that a has $s+1$ continuous derivatives. Then (17.7) is obtained if we differentiate the equality $\dfrac{a(t)}{t} = \displaystyle\int_0^1 a'(\eta t)\,d\eta$ s times and then integrate by parts. □

Let us now pass to the proof of Theorem 17.1. Applying (17.3)–(17.5) to the function $A_s * H$ with Fourier–Stieltjes transform $a_s(t)h(t)$, we see that

$$\|A_s * H\| \leqslant \frac{1}{\pi}\sup_x \left|\int_{-\infty}^{\infty} e^{-itx}h(t)\left[\int_0^1 \frac{a^{(s)}(t) - a^{(s)}(\eta t)}{t}\,d\eta^s\right]dt\right|. \tag{17.8}$$

We verify that we can change the order of integration on the right-hand side; this would follow from the fact that

$$\delta_0(\varepsilon,\varepsilon') \equiv \int_{\varepsilon'<|t|<\varepsilon} e^{-itx}h(t)\left[\int_0^1 \frac{a^{(s)}(\eta t)}{t}\,d\eta^s\right]dt \to 0, \qquad 0<\varepsilon'<\varepsilon, \quad \varepsilon\to 0,$$

and

$$\delta_1(T,T') \equiv \int_{T<|t|<T'} e^{-itx}h(t)\left[\int_0^1 \frac{a^{(s)}(\eta t)}{t}\,d\eta^s\right]dt \to 0, \qquad T<T', \quad T\to\infty.$$

For any $\eta \in (0,1)$ the function $t \mapsto a^{(s)}(\eta t)h(t)$ is, up to the factor i^s, the Fourier–Stieltjes transform of the convolution $V_\eta(x) = B(x/\eta) * H(x)$, where $B(x) = \displaystyle\int_{-\infty}^{x} y^s\,dA(y)$. Hence, introducing the function $\psi(t) = \displaystyle\int_0^t \frac{\sin u}{u}\,du$ and letting $V = V_1$, we obtain

$$\begin{aligned}
\delta_1(T,T') &= \int_0^1\left[\int_{T<|t|<T'} e^{-itx}h(t)\frac{a^{(s)}(\eta t)}{t}\,dt\right]d\eta^s\\
&= \int_0^1\left[\int_{-\infty}^{\infty}\int_{T<|t|<T'} \frac{e^{it(y-x)}}{t}\,dt\,dV_\eta(y)\right]d\eta^s\\
&= 2i\int_0^1\int_{-\infty}^{\infty}\bigl(\psi(T'(y-x)) - \psi(T(y-x))\bigr)\,dV_\eta(y)\,d\eta^s\\
&= 2i\int_0^1\int_{-\infty}^{\infty} u(y,\eta)\,dV(y)\,d\eta^s,
\end{aligned}$$

where

$$u(y,\eta) = \psi\big(T'(\eta y - x)\big) - \psi\big(T(\eta y - x)\big).$$

Since $|\psi| \leqslant C$, $\psi(t) \to \pi/2$ as $t \to \infty$, and $\psi(t) \to -\pi/2$ as $t \to -\infty$, the modulus of the function u is bounded by the absolute constant $2C$ uniformly over all $T' > T > 0$, and $u(y,\eta) \to 0$ as $T \to \infty$ for any y and η. Consequently, by the Lebesgue dominated convergence theorem we have $\delta_1(T,T') \to 0$. A similar argument proves that also $\delta_0(\varepsilon,\varepsilon') \to 0$ as $\varepsilon \to 0$.

Finally, changing the order of integration, we can conclude that the expression under the supremum sign on the right-hand side of (17.8) does not exceed the quantity

$$\left|\int_{-\infty}^{\infty} e^{-itx} h(t) \frac{a^{(s)}(t)}{t}\,dt\right| + \sup_{0<\eta<1}\left|\int_{-\infty}^{\infty} e^{-itx} h(t) \frac{a^{(s)}(\eta t)}{t}\,dt\right|,$$

which implies the required inequality (17.1). Theorem 17.1 is proved. □

Remark 17.3. For $s = 0$ Theorem 17.1 remains valid, and without the condition $A(\infty) = 0$. Furthermore, the inequality (17.1) can be improved. Indeed, applying (17.3) and then the left-hand inequality in (17.4) with $B = A * H$ and $b = ah$, we obtain

$$\|A * H\| \leqslant \frac{1}{\pi} \sup_x \left|\int_{-\infty}^{\infty} e^{-itx} \frac{a(t)}{t} h(t)\,dt\right|. \tag{17.9}$$

If $A(\infty) = 0$, then the factor $1/\pi$ can be replaced by $1/(2\pi)$. In this case $B(\infty) = 0$, and thus (17.3) implies the identity

$$\frac{1}{\pi} \sup_x \left|\int_{-\infty}^{\infty} \frac{e^{-itx}}{t}\, b(t)\,dt\right| = 2\|B\|.$$

Returning to the 'non-smoothed' variant (17.2), we make another remark. Since up to the factor i^s the derivative $a^{(s)}$ is the Fourier–Stieltjes transform of the function $B(x) = \int_{-\infty}^x y^s\,dA(y)$, the modulus of the integral on the right-hand side of (17.2) coincides by (17.3) with

$$\pi\left|\int_{-\infty}^{\infty} \big(1_{\{y<x\}} - 1_{\{y>x\}}\big) y^s\,dA(y)\right|.$$

Therefore, in view of the right-hand inequality in (17.4), (17.2) implies the estimate

$$\sup_x |x^s A(x)| \leqslant 6 \sup_x \left|\int_{-\infty}^{x} y^s\,dA(y)\right|,$$

which does not involve Fourier–Stieltjes transforms. By (17.4), this estimate is equivalent to (17.2) up to an absolute multiplicative constant.

18. Non-uniform estimates for distribution functions

Now we return to Lemma 16.1 and apply Theorem 17.1 and the inequality (17.9) with $A = F - G$ and the smoothing distribution function $H(Tx)$ (instead of H). Then we arrive at a non-uniform estimate, that is, an estimate for the distance

$$\rho_s(F,G) = \sup_x (1 + |x|^s)|F(x) - G(x)|$$

in terms of the Fourier–Stieltjes transforms

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x) \quad \text{and} \quad g(t) = \int_{-\infty}^{\infty} e^{itx}\, dG(x), \qquad t \in \mathbb{R}.$$

We assume that the distribution function F and the function G of bounded variation have finite absolute moments of integer order $s \geqslant 1$ and that $G(-\infty) = 0$ and $G(\infty) = 1$.

Theorem 18.1. *If G is differentiable and satisfies the inequality*

$$\sup_x (1 + |x|^s)|G'(x)| \leqslant L,$$

then for any characteristic function h and all $T \geqslant 1$

$$c_s \rho_s(F, G) \leqslant \sup_{x\in\mathbb{R}} \left| \int_{-\infty}^{\infty} e^{-itx} \frac{f(t) - g(t)}{t} h\left(\frac{t}{T}\right) dt \right| + \sup_{x\in\mathbb{R},\ \kappa\geqslant 1} \left| \int_{-\infty}^{\infty} e^{-itx} \frac{f^{(s)}(t) - g^{(s)}(t)}{t} h\left(\frac{\kappa t}{T}\right) dt \right| + \frac{L}{T} \tag{18.1}$$

with some constant $c_s > 0$ depending on s and h.

In particular, using the canonical kernel $h(t) = (1 - |t|)_+$ and taking into account that $\kappa \geqslant 1$ under the second supremum sign in (18.1), we get that

$$c_s \rho_s(F, G) \leqslant \int_{-T}^{T} \left| \frac{f(t) - g(t)}{t} \right| dt + \int_{-T}^{T} \left| \frac{f^{(s)}(t) - g^{(s)}(t)}{t} \right| dt + \frac{L}{T}. \tag{18.2}$$

Without using the second identity in Lemma 17.2, we would arrive at a similar estimate

$$c_s \rho_s(F, G) \leqslant \int_{-T}^{T} \left| \frac{f(t) - g(t)}{t} \right| dt + \int_{-T}^{T} \left| \frac{d^s}{dt^s}\, \frac{f(t) - g(t)}{t} \right| dt + \frac{L}{T}, \tag{18.3}$$

which the reader can find, for example, in Petrov's book [3].

It makes sense to use (18.2) and (18.3) in the case when F and G have finite absolute moments of order $s + 1$, and

$$\int_{-\infty}^{\infty} x^p\, dF(x) = \int_{-\infty}^{\infty} x^p\, dG(x), \qquad p = 1, \dots, s.$$

In this case $|f^{(s)}(t) - g^{(s)}(t)| = O(|t|)$ as $t \to 0$, so that the integrals on the right-hand sides of these inequalities are finite. But if only the moments of F and G of orders $\leqslant s - 1$ are equal, then we need additional arguments, which, furthermore, can also be based on the smoothing inequality (18.1).

We assume, for example, that for the derivatives of f and g of order s there is a decomposition of the form

$$f^{(s)}(t) - g^{(s)}(t) = u(t) + v(t), \tag{18.4}$$

where v is the Fourier–Stieltjes transform of a function V of bounded variation such that $V(-\infty) = 0$ but not necessarily $V(\infty) = 0$. If it is known that the norm $\|V\|$ is small (preferably of order $1/T$) and $u(t) = tw(t)O(1/T)$ on the interval $[-T, T]$ for some integrable function w, then the following application of Theorem 18.1 can be useful.

Corollary 18.2. *Let the conditions of Theorem 18.1 hold, and assume (18.4). Then*

$$c_s \rho_s(F, G) \leqslant \int_{-T}^{T} \left| \frac{f(t) - g(t)}{t} \right| dt + \int_{-T}^{T} \left| \frac{u(t)}{t} \right| dt + \|V\| + \frac{L}{T} \tag{18.5}$$

for all $T \geqslant 1$, where $c_s > 0$ is a constant depending only on s.

To derive this inequality on the basis of (18.1), we only need to estimate the integrals

$$I = \int_{-\infty}^{\infty} e^{-itx} \frac{v(t)}{t} h\left(\frac{\kappa t}{T}\right) dt, \quad \text{where} \quad h(t) = (1 - |t|)_+,$$

uniformly over all x and κ. Let us denote the corresponding distribution function by H and put $H_q(x) = H(qx)$. Then by (17.3) and the upper estimate in (17.4) we have

$$|I| = \pi \left| \int_{-\infty}^{\infty} \bigl(1_{\{y<x\}} - 1_{\{y>x\}}\bigr)\, d(V * H_{T/\kappa})(y) \right| \leqslant 3\pi \|V * H_{T/\kappa}\| \leqslant 3\pi \|V\|$$

(the last inequality in this chain holds because the convolution with any distribution function does not increase the L^∞-norm of a given function of bounded variation). As a result, we obtain (18.5).

To illustrate the decomposition (18.4), we assume that F is a convolution of probability distributions F_k $(1 \leqslant k \leqslant n)$ with characteristic functions f_k such that the F_k have zero means, variances σ_k^2 with $\sum_{k=1}^n \sigma_k^2 = 1$, and finite moments $\beta_{s,k}$ of order $s \geqslant 3$. Then F has the characteristic function $f(t) = f_1(t) \cdots f_n(t)$. As G we can take the standard normal distribution function Φ or, better, the 'corrected' normal function from the Edgeworth expansion of order s (which is not necessarily monotone). Then we can let

$$v(t) = \sum_{k=1}^{n} f_1(t) \cdots f_{k-1}(t)\, f_k^{(s)}(t)\, f_{k+1}(t) \cdots f_n(t).$$

Obviously, the corresponding function V of bounded variation has norm

$$\|V\| \leqslant \sum_{k=1}^{n} \sup_x \left| \int_{-\infty}^{x} y^s\, dF_k(y) \right| \leqslant \sum_{k=1}^{n} \beta_{s,k} \equiv L_s,$$

that is, this norm can be estimated by the Lyapunov fraction L_s of order s. It can be shown that for $s = 3$ the second component $u(t) = f^{(3)}(t) - g^{(3)}(t) - v(t)$ of the decomposition (18.4) can be estimated by a similar fraction with coefficient

$O(|t|e^{-t^2/6})$ on the interval $|t| \leqslant T$, where $T = 1/L_3$. Thus, in this case the use of (18.5) yields a well-known estimate of the form

$$\rho_3(F, \Phi) \leqslant cL_3 \tag{18.6}$$

with an absolute constant c. (Under additional assumptions about the behaviour of $f_k(t)$ for large t, similar estimates $\rho_s(F, G) \leqslant c_s L_s$ are valid for $s \geqslant 4$.)

For non-uniform estimates in the central limit theorem, Pinelis [43] (see also [44]) recently called attention to the possibility of such an approach connected with the selection of the 'bad' component $v(t)$ in the decomposition (18.4). He proposed a new proof of (18.6) for convolution powers (distributions of sums of independent identically distributed random variables), based on Bohman–Prawitz–Vaaler-type smoothing inequalities

$$\begin{aligned} F(x) &\leqslant \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-itx}}{-it} f(t) h\biggl(\frac{t}{T}\biggr)\, dt, \\ F(x-) &\geqslant \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-itx}}{-it} f(t) h\biggl(-\frac{t}{T}\biggr)\, dt \end{aligned} \qquad (x \in \mathbb{R},\ T > 0), \tag{18.7}$$

which are valid in the class of all distribution functions F with characteristic functions f. In such inequalities the functions h (the so-called smoothing kernels), which are Fourier–Stieltjes transforms of signed measures, must have a special structure. In applications, they are chosen to have support in the interval $[-1, 1]$, so that the integrals in (18.7) are taken over the interval $[-T, T]$. For example, the Prawitz kernel is given by the formula

$$h(t) = (1 - |t|)\,\pi t \cot(\pi t) + |t| - i(1 - |t|)\pi t, \qquad |t| < 1$$

(see [45]). Following Bohman [46], Pinelis described a broad class of functions h satisfying (18.7). If F has a finite absolute moment of some order s and h has s continuous derivatives, then it was shown in [43] that for all $x \geqslant 0$

$$\begin{aligned} x^s(1 - F(x)) &\leqslant \frac{(-i)^s}{2\pi}\int_{-\infty}^{\infty} e^{-itx} \frac{d^s}{dt^s}\, \frac{f(t)h(-t/T) - 1}{it}\, dt, \\ x^s(1 - F(x-)) &\geqslant \frac{(-i)^s}{2\pi}\int_{-\infty}^{\infty} e^{-itx} \frac{d^s}{dt^s}\, \frac{f(t)h(t/T) - 1}{it}\, dt, \end{aligned}$$

which can be regarded as an analogue of Theorem 18.1. Other interesting applications of inequalities like (18.7) and related extremum problems in Fourier analysis were discussed by Vaaler in [47].

We note that the use of inequalities like (18.7) enables one to avoid the problem of comparing the distances between the initial and the smoothed probability distributions and can lead to the improvement of a number of results (for example, on the choice of absolute constants). On the other hand, the scope of application of (18.7) is limited to the class of monotone functions G playing the role of approximations of F, in contrast to more general inequalities like (18.1) and (18.5).

19. Lower estimates for the Kolmogorov distance

In conclusion, we consider the opposite problem, that is, estimating the Kolmogorov distance between distribution functions from below in terms of the corresponding characteristic functions. Since approximations of a given distribution function by functions which are not necessarily monotone are of interest in some problems, it makes sense to consider a wider class of functions A of bounded variation and to estimate the L^∞-norm $\|A\|$ from below in terms of the Fourier–Stieltjes transforms

$$a(t) = \int_{-\infty}^{\infty} e^{itx}\, dA(x).$$

We assume that $A(-\infty) = A(\infty) = 0$.

First we present one popular estimate (see [13], [48], [3]).

Theorem 19.1.

$$\|A\| \geqslant \frac{1}{2\sqrt{2\pi}} \biggl| \int_{-\infty}^{\infty} a(t) e^{-t^2/2}\, dt \biggr|. \tag{19.1}$$

If the behavior of the function $a(t)$ is known only in a neighbourhood of zero, then the following theorem (which has apparently not been mentioned in the literature) may be preferable.

Theorem 19.2. *For any $T > 0$*

$$\|A\| \geqslant \frac{1}{3T} \biggl| \int_0^T a(t) \Bigl(1 - \frac{t}{T}\Bigr)\, dt \biggr|. \tag{19.2}$$

A standard approach to estimates of this kind is based on Plancherel's theorem, that is, on the identity

$$\int_{-\infty}^{\infty} v(t) w(t)\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathscr{P}v(x) \overline{\mathscr{P}\overline{w}}(x)\, dx, \qquad v, w \in L^2(\mathbb{R}), \tag{19.3}$$

where $\mathscr{P}v = \widehat{v}$ denotes the Fourier transform of the function v. If the function $a(t)/t$ is integrable, then by the inversion formula

$$A(x) = -\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{a(t)}{it}\, dt.$$

In other words $A(-x)$ is the Fourier transform of the function $v(t) = -\dfrac{1}{2\pi i}\, \dfrac{u(t)}{t}$. If also $v \in L^2$, then (19.3) assumes the form

$$\int_{-\infty}^{\infty} \frac{a(t)}{t}\, w(t)\, dt = -i \int_{-\infty}^{\infty} A(-x) \overline{\mathscr{P}\overline{w}}(x)\, dx = -i \int_{-\infty}^{\infty} A(x) \mathscr{P}w(x)\, dx.$$

This immediately implies the general estimate

$$\biggl| \int_{-\infty}^{\infty} \frac{a(t)}{t}\, w(t)\, dt \biggr| \leqslant \|A\| \int_{-\infty}^{\infty} |\mathscr{P}w(x)|\, dx. \tag{19.4}$$

For example, the particular case $w(t) = te^{-t^2/2}$ yields (19.1). Additional assumptions about $a(t)$ in this inequality, as well as in (19.2), can easily be dropped. To derive (19.2), we apply (19.4) to $w(t) = (t/T)(1 - t/T)_+ 1_{(0,\infty)}(t)$. We first put $T = 1$. Then for all $x \neq 0$ we have

$$\mathscr{P}w(x) = \int_0^1 e^{itx} t(1-t)\,dt = \frac{-e^{ix} - 1}{x^2} + \frac{2(e^{ix} - 1)}{ix^3} = \frac{q(x)}{x^3}\,,$$

where $q(x) = -xe^{ix} - 2i\,e^{ix} - x + 2i$. This implies the inequalities $|q(x)| \leqslant 2|x| + 4$ and $|\mathscr{P}w(x)| \leqslant (2|x|+4)/|x|^3$. On the other hand,

$$|\mathscr{P}w(x)| \leqslant \int_0^1 t(1-t)\,dt = \frac{1}{6}$$

and hence

$$\begin{aligned}\int_{-\infty}^{\infty} |\mathscr{P}w(x)|\,dx &= 2\int_0^4 |\mathscr{P}w(x)|\,dx + 2\int_4^{\infty} |\mathscr{P}w(x)|\,dx \\ &\leqslant \frac{4}{3} + 2\int_4^{\infty} \frac{2x+4}{x^3}\,dx = \frac{17}{6} < 3.\end{aligned}$$

In the general case, $\mathscr{P}w_T(x) = T \cdot (\mathscr{P}w)(Tx)$ for $w_T(x) = w(t/T)$, so that

$$\int_{-\infty}^{\infty} |\mathscr{P}w_T(x)|\,dx = \int_{-\infty}^{\infty} |\mathscr{P}w(x)|\,dx < 3.$$

Thus, the use of $w = w_T$ in (19.4) proves (19.2).

We also note that, applying (19.4) to the different function

$$w(t) = \frac{t}{T}\left(1 - \frac{|t|}{T}\right)_+,$$

we would obtain the analogous inequality

$$\|A\| \geqslant \frac{1}{3.5T}\left|\int_{-T}^{T} a(t)\left(1 - \frac{|t|}{T}\right)dt\right|.$$

20. Estimates in the central limit theorem

Let $F_n(x) = \mathsf{P}\{Z_n \leqslant x\}$ be the distribution function of the normalized sum

$$Z_n = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$$

of independent identically distributed random variables $X_1, \ldots, X_n$ such that $\mathsf{E}X_1 = 0$ and $\mathsf{E}X_1^2 = 1$. According to the central limit theorem, as n increases, $F_n(x)$ converges to the standard normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy.$$

How close F_n is to Φ in a given metric d depends on the initial distribution of the sample, that is, on the distribution of X_1. Nevertheless, for the classical metrics giving weak convergence the rate of convergence of the distance $d(F_n,\Phi)$ to zero is at least $c/\sqrt{n}$ under sufficiently broad assumptions.

To establish results of this kind, it suffices to compare the characteristic functions

$$f_n(t)=\int_{-\infty}^{\infty}e^{itx}\,dF_n(x)=f_1\biggl(\frac{t}{\sqrt{n}}\biggr)^n \quad\text{and}\quad g(t)=\int_{-\infty}^{\infty}e^{itx}\,d\Phi(x)=e^{-t^2/2}$$

and use the corresponding smoothing inequalities. If the third absolute moment $\beta_3=\mathsf{E}|X_1|^3$ is finite, then it is not difficult to expand $f_1(t)$ by Taylor's formula in a neighbourhood of zero (up to the cubic term) and, as a consequence, to obtain an estimate of the form

$$|f_n(t)-g(t)|\leqslant c\frac{\beta_3}{\sqrt{n}}|t|^3e^{-t^2/4},\qquad |t|\leqslant\frac{\sqrt{n}}{\beta_3}\,, \tag{20.1}$$

with some absolute constant $c>0$. A similar relation also holds for the first three derivatives. More precisely, in the same interval

$$|f_n^{(s)}(t)-g^{(s)}(t)|\leqslant c\frac{\beta_3}{\sqrt{n}}|t|^{3-s}e^{-t^2/4},\qquad s=0,1,2,3. \tag{20.2}$$

20.1. Kolmogorov and Lévy distances in the L^p-metric. Use of the estimate (20.1) in Theorem 2.1 with $T=\sqrt{n}/\beta_3$ yields the classical Berry–Esseen theorem for the Kolmogorov distance.

Theorem 20.1. *The inequality*

$$\rho(F_n,\Phi)\leqslant c\frac{\beta_3}{\sqrt{n}} \tag{20.3}$$

holds with some absolute constant $c>0$.

The value of the best constant c in this inequality is unknown. It is known only that $0.4097<c<0.4690$ (see [49]). A similar statement also holds for the Lévy distance, since $L(F,\Phi)\leqslant\rho(F,\Phi)\leqslant(1+1/\sqrt{2\pi})L(F,\Phi)$ in the general case.

But if the third absolute moment β_3 is infinite, then the rate of convergence to the normal law can be arbitrarily slow. As Matskyavichyus [48] showed using the lower estimate (19.1), for any numerical sequence $\varepsilon_n\to 0$ (as $n\to\infty$) one can choose a common distribution of random variables $X_1,X_2,\dots$ with zero mean and unit variance such that for all sufficiently large n

$$\rho(F_n,\Phi)\geqslant\varepsilon_n.$$

On the other hand, (20.3) can be strengthened in terms of the Kolmogorov distance with a polynomial weight. If in the Berry–Esseen-type inequality (18.2) we use (20.2) with $s=2$ and the same parameter $T=\sqrt{n}/\beta_3$, then an estimate due to Meshalkin and Rogozin [50] is obtained:

$$|F_n(x)-\Phi(x)|\leqslant c\frac{\beta_3}{(1+x^2)\sqrt{n}}\,. \tag{20.4}$$

In fact, there is an even stronger statement, proved by Nagaev using exponential estimates and additional constructions like truncation (see [51], [52], [2]).

Theorem 20.2. *For all $x \in \mathbb{R}$ the inequality*

$$|F_n(x) - \Phi(x)| \leqslant c\frac{\beta_3}{(1+|x|^3)\sqrt{n}} \tag{20.5}$$

holds with some absolute constant $c > 0$.

Nagaev's approach has been further developed in many investigations, including [53]–[57], where ways of refining the constant c in (20.5) were also studied. As we mentioned above in our discussion of Pinelis' alternative approach [43], [44], this inequality can also be obtained with the help of Corollary 18.2 with the function

$$v(t) = \frac{1}{\sqrt{n}} f_1\biggl(\frac{t}{\sqrt{n}}\biggr)^{n-1} f_1'''\biggl(\frac{t}{\sqrt{n}}\biggr),$$

which plays the role of the singular ('bad') component of the decomposition (18.4) for the third derivative $f_n'''(t)$.

We also note that the inequalities (20.4) and (20.5) immediately yield upper estimates for the distance in the L^p-metric:

$$\|F_n - \Phi\|_p \leqslant c\frac{\beta_3}{\sqrt{n}}\,, \qquad p \geqslant 1.$$

In particular, for $p = 1$ we arrive at Esseen's inequality for the mean distance, which can also be derived on the basis of Theorem 8.2 (or Corollary 8.3).

20.2. Lévy–Prokhorov distance. The non-uniform estimate (20.5) does not suffice for studying the rate of convergence in other metrics on its basis, and thus other smoothing inequalities are needed. If we use (20.2) with $s = 1$ in Theorem 6.1 with $T = \sqrt{n}/\beta_3$, then we arrive at a theorem of Yurinskii [16] which strengthens Theorem 20.1.

Theorem 20.3. *The inequality*

$$\pi(F_n, \Phi) \leqslant c\frac{\beta_3}{\sqrt{n}}$$

holds with an absolute constant $c > 0$.

This also holds for sums of identically distributed random vectors in $\mathbb{R}^k$ with a constant c depending only on the dimension k.

20.3. Zolotarev distances. For convenience we write $\zeta_s(X, Y)$ instead of $\zeta_s(F, G)$ if the random variables X and Y have distributions F and G, respectively $(s = 1, 2, \dots)$. We mention two simple but important properties of these ideal metrics:

1) $\zeta_s(\lambda X, \lambda Y) = \lambda^s \zeta_s(X, Y)$ for all $\lambda \geqslant 0$ (homogeneity of order s);

2) $\zeta_s(X + X', Y + Y') \leqslant \zeta_s(X, Y) + \zeta_s(X', Y')$ if the random variables X and X' are independent and Y and Y' are also independent (semi-additivity).

Under the assumptions of the central limit theorem with finite third moment, we immediately obtain the following theorem of Zolotarev.

Theorem 20.4.

$$\zeta_3(F_n,\Phi) \leqslant \frac{1}{\sqrt{n}}\zeta_3(F_1,\Phi).$$

Rather interestingly, the desired rate of order $1/\sqrt{n}$ is ensured by the above properties 1) and 2) of the metric ζ_3. In a more general case the inequality

$$\zeta_s(F_n,\Phi) \leqslant \frac{1}{n^{(s-2)/2}}\zeta_s(F_1,\Phi)$$

holds, though for $s \geqslant 4$ it makes sense only under the additional assumptions that $\mathsf{E}|X_1|^s < \infty$ and $\mathsf{E}X_1^p = \mathsf{E}Z^p$ ($p = 3,\dots,s-1$), where Z is a random variable with standard normal distribution. In particular, if $\beta_4 = \mathsf{E}X_1^4 < \infty$ and $\mathsf{E}X_1^3 = 0$ (for example, if the initial distribution F_1 is symmetric with respect to the origin), then

$$\zeta_4(F_n,\Phi) = O\biggl(\frac{1}{n}\biggr).$$

However, this result can be significantly strengthened in terms of the metric ζ_2. Assuming finiteness of the fourth moment, one can expand $f_1(t)$ by Taylor's formula in powers of t in a neighbourhood of zero, with a remainder of the form $\beta_4 t^4$, and, as a consequence, obtain an improvement of (20.2), namely,

$$|f_n^{(s)}(t) - \widetilde{g}^{(s)}(t)| \leqslant c\frac{\beta_4}{n}|t|^{4-s}e^{-t^2/4}, \qquad |t| \leqslant \frac{\sqrt{n}}{\beta_3}, \quad s = 0,1,2,3,4. \tag{20.6}$$

Here

$$\widetilde{g}(t) = e^{-t^2/2}\biggl(1 + \mathsf{E}X_1^3\frac{(it)^3}{3!}\,\frac{1}{\sqrt{n}}\biggr)$$

is the Fourier–Stieltjes transform of the 'corrected' Gaussian distribution. In particular, $\widetilde{g}(t) = g(t)$ if $\mathsf{E}X_1^3 = 0$.

Applying (20.6) with $s = 1$ in Theorem 9.1, we can prove the following inequality, which apparently cannot be found in the literature.

Theorem 20.5. *If $\mathsf{E}X_1^3 = 0$ and $\beta_4 = \mathsf{E}X_1^4 < \infty$, then the inequality*

$$\zeta_2(F_n,\Phi) \leqslant c\frac{\beta_4}{n}$$

holds with some absolute constant $c > 0$.

Since $\zeta_1^2 \leqslant c'\zeta_2$, this estimate agrees with the estimate $\zeta_1(F_n,\Phi) \leqslant c\beta_3/\sqrt{n}$ for the mean distance (asymptotically with respect to n).

20.4. Kantorovich distances. The best result known in the problem of the rate of convergence for Kantorovich transport metrics in the central limit theorem is the following assertion, which was proved by Rio [35] using the relation (10.3). We give it in the case of identically distributed summands, as in the previous theorems (see also [58]).

Theorem 20.6. *If $\beta_{p+2} = \mathsf{E}|X_1|^{p+2} < \infty$ for $1 \leqslant p \leqslant 2$, then*

$$W_p(F_n, \Phi) \leqslant c_p \frac{\beta_{p+2}^{1/p}}{\sqrt{n}}\,, \tag{20.7}$$

where $c_p > 0$ is a constant depending only on p.

For $p = 1$ we return to a known estimate for the mean distance, and for $p = 2$ we arrive at an estimate for the quadratic Kantorovich distance:

$$W_2(F_n, \Phi) \leqslant c\frac{\sqrt{\beta_4}}{\sqrt{n}}\,. \tag{20.8}$$

Interestingly, the finiteness of the fourth moment is essential for attaining the standard rate $1/\sqrt{n}$ in the metric W_2. Another approach (the so-called entropy approach) to the derivation of (20.8) was proposed in [59]. The inequality (20.8) can also be obtained using (20.2) if we invoke the smoothing inequality in Theorem 11.1 with $g_\alpha = \widetilde{g}$ and $T = \sqrt{n}/\beta_3$. As for values $p > 2$, in this case the inequality (20.7) also remains valid. Using Edgeworth expansions, the author recently proved such an inequality in [60].

In conclusion, we note that all the above estimates remain valid for sums of non-identically distributed summands after corresponding modifications in terms of Lyapunov fractions. There are also a variety of similar results for 'strong' metrics such as the distance in variation, the Kullback–Leibler divergence, and so on. However, moment conditions alone are not sufficient for convergence of F_n to Φ with any rate in strong metrics (see [61]).

The author is grateful to Irina Shevtsova and Andrei Zaitsev for valuable remarks and stimulating discussions.

Bibliography

[1] W. Feller, *An introduction to probability theory and its applications*, vol. II, John Wiley & Sons, Inc., New York–London–Sydney 1966, xviii+636 pp.

[2] В. В. Петров, *Суммы независимых случайных величин*, Наука, М. 1972, 416 с.; English transl., V. V. Petrov, *Sums of independent random variables*, Ergeb. Math. Grenzgeb., vol. 82, Springer-Verlag, New York–Heidelberg 1975, x+346 pp.

[3] В. В. Петров, *Предельные теоремы для сумм независимых случайных величин*, Наука, М. 1987, 318 с. [V. V. Petrov, *Limit theorems for sums of independent random variables*, Nauka, Moscow 1987, 318 pp.]

[4] C.-G. Esseen, "Fourier analysis of distribution functions. A mathematical study of the Laplace–Gaussian law", *Acta Math.* **77** (1945), 1–125.

[5] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates", *Trans. Amer. Math. Soc.* **49** (1941), 122–136.

[6] А. С. Файнлейб, "Обобщение неравенства Эссеена и его применение в вероятностной теории чисел", *Изв. АН СССР. Сер. матем.* **32**:4 (1968), 859–879; English transl., A. S. Fainleib, "A generalization of Esseen's inequality and its application in probabilistic number theory", *Math. USSR-Izv.* **2**:4 (1968), 821–844.

[7] V. Bentkus and F. Götze, “Optimal rates of convergence in the CLT for quadratic forms”, *Ann. Probab.* **24**:1 (1996), 466–490.

[8] H. Bohman, “Approximate Fourier analysis of distribution functions”, *Ark. Mat.* **4**:2 (1961), 99–157.

[9] В. М. Золотарев, “О свойствах и связях некоторых типов метрик”, *Исследования по математической статистике*. 3, Зап. науч. сем. ЛОМИ, **87**, Изд-во “Наука”, Ленинград. отд., Л. 1979, с. 18–35; English transl., V. M. Zolotarev, “Properties of and relations among certain types of metrics”, *J. Soviet Math.* **17**:6 (1981), 2218–2232.

[10] Ю. В. Линник, И. В. Островский, *Разложения случайных величин и векторов*, Наука, М. 1972, 479 с.; English transl., Yu. V. Linnik and I. V. Ostrovskii, *Decomposition of random variables and vectors*, Transl. Math. Monogr., vol. 48, Amer. Math. Soc., Providence, RI 1977, ix+380 pp.

[11] В. М. Золотарев, “Оценки различия распределений в метрике Леви”, *Сборник статей*. I, Посвящается академику Ивану Матвеевичу Виноградову к его восьмидесятилетию, Тр. МИАН СССР, **112**, 1971, с. 224–231; English transl., V. M. Zolotarev, “Estimates of the difference between distributions in the Lévy metric”, *Proc. Steklov Inst. Math.* **112** (1971), 232–240.

[12] А. Ю. Зайцев, “О логарифмическом множителе в неравенствах сглаживания для расстояний Леви и Леви–Прохорова”, *Теория вероятн. и ее примен.* **31**:4 (1986), 782–784; English transl., A. Yu. Zaitsev, “On the logarithmic factor in smoothing inequalities for Lévi and Lévi–Prohorov distances”, *Theory Probab. Appl.* **31**:4 (1987), 691–693.

[13] И. А. Ибрагимов, Ю. В. Линник, *Независимые и стационарно связанные величины*, Наука, М. 1965, 524 с.; English transl., I. A. Ibragimov and Yu. V. Linnik, *Independent and stationary sequences of random variables*, Wolters-Noordhoff Publishing Company, Groningen 1971, 443 pp.

[14] S. G. Bobkov, G. P. Chistyakov, and F. Götze, “Berry–Esseen bounds in the entropic central limit theorem”, *Probab. Theory Related Fields* **159**:3-4 (2014), 435–478.

[15] Ю. В. Прохоров, “Сходимость случайных процессов и предельные теоремы теории вероятностей”, *Теория вероятн. и ее примен.* **1**:2 (1956), 177–238; English transl., Yu. V. Prokhorov, “Convergence of random processes and limit theorems in probability theory”, *Theory Probab. Appl.* **1**:2 (1956), 157–214.

[16] В. В. Юринский, “Неравенство сглаживания для оценок расстояния Леви–Прохорова”, *Теория вероятн. и ее примен.* **20**:1 (1975), 3–12; English transl., V. V. Yurinskii, “A smoothing inequality for estimates of the Lévy–Prokhorov distance”, *Theory Probab. Appl.* **20**:1 (1975), 1–10.

[17] В. А. Абрамов, “Оценка расстояния Леви–Прохорова”, *Теория вероятн. и ее примен.* **21**:2 (1976), 406–410; English transl., V. A. Abramov, “Estimates for the Lévy–Prokhorov distance”, *Theory Probab. Appl.* **21**:2 (1976), 396–400.

[18] А. Ю. Зайцев, “Оценки расстояния Леви–Прохорова в терминах характеристических функций и некоторые их применения”, *Проблемы теории вероятностных распределений*. VII, Зап. науч. сем. ЛОМИ, **119**, Изд-во “Наука”, Ленинград. отд., Л. 1982, с. 108–127; English transl., A. Yu. Zaitsev, “Estimates for the Levy–Prokhorov distance in terms of characteristic functions and some of their applications”, *J. Soviet Math.* **27** (1984), 3070–3083.

[19] Т. В. Арак, А. Ю. Зайцев, “Равномерные предельные теоремы для сумм независимых случайных величин”, Тр. МИАН СССР, **174**, Изд-во “Наука”, Ленинград. отд., Л. 1986, с. 3–214; English transl., T. V. Arak

and A. Yu. Zaitsev, "Uniform limit theorems for sums of independent random variables", *Proc. Steklov Inst. Math.* **174** (1988), 1–222.

[20] W. Beckner, "Inequalities in Fourier analysis", *Ann. of Math.* (2) **102**:1 (1975), 159–182.

[21] C.-G. Esseen, "On mean central limit theorems", *Kungl. Tekn. Högsk. Handl. Stockholm*, 1958, № 121, 30 с.

[22] В. М. Золотарев, "Идеальные метрики в проблеме аппроксимации распределений сумм независимых случайных величин", *Теория вероятн. и ее примен.* **22**:3 (1977), 449–465; English transl., V. M. Zolotarev, "Ideal metrics in the problem of approximating distributions of sums of independent random variables", *Theory Probab. Appl.* **22**:3 (1978), 433–449.

[23] V. M. Zolotarev, *Modern theory of summation of random variables*, Mod. Probab. Stat., VSP, Utrecht 1997, x+412 pp.

[24] В. В. Сенатов, "Одна оценка метрики Леви–Прохорова", *Теория вероятн. и ее примен.* **29**:1 (1984), 108–113; English transl., V. V. Senatov, "An estimate of the Lévy–Prohorov metric", *Theory Probab. Appl.* **29**:1 (1985), 109–114.

[25] V. M. Zolotarev, "Ideal metrics in the problems of probability theory and mathematical statistics", *Austral. J. Statist.* **21**:3 (1979), 193–208.

[26] A. M. Vershik, "Long history of the Monge–Kantorovich transportation problem", *Math. Intelligencer* **35**:4 (2013), 1–9.

[27] Л. В. Канторович, "О перемещении масс", *Докл. АН СССР* **37**:7-8 (1942), 227–229; *Теория представлений, динамические системы*. XI, Специальный выпуск, Зап. науч. сем. ПОМИ, **312**, ПОМИ, СПб. 2004, с. 11–14; English transl., L. V. Kantorovich, "On the translocation of masses", *J. Math. Sci.* (*N. Y.*) **133**:4 (2006), 1381–1382.

[28] Л. В. Канторович, "Об одной проблеме Монжа", в ст. "Заседания Московского математического общества (резюме докладов)", *УМН* **3**:2(24) (1948), 225–226; *Теория представлений, динамические системы*. XI, Специальный выпуск, Зап. науч. сем. ПОМИ, **312**, ПОМИ, СПб. 2004, с. 15–16; English transl., L. V. Kantorovich, "On a problem of Monge", *J. Math. Sci.* (*N. Y.*) **133**:4 (2006), 1383.

[29] С. Т. Рачев, "Задача Монжа–Канторовича о перемещении масс и ее применение в стохастике", *Теория вероятн. и ее примен.* **29**:4 (1984), 625–653; English transl., S. T. Rachev, "The Monge–Kantorovich mass transference problem and its stochastic applications", *Theory Probab. Appl.* **29**:4 (1985), 647–676.

[30] C. Villani, *Topics in optimal transportation*, Grad. Stud. Math., vol. 58, Amer. Math. Soc., Providence, RI 2003, xvi+370 pp.

[31] В. И. Богачев, А. В. Колесников, "Задача Монжа–Канторовича: достижения, связи и перспективы", *УМН* **67**:5(407) (2012), 3–110; English transl., V. I. Bogachev and A. V. Kolesnikov, "The Monge–Kantorovich problem: achievements, connections, and perspectives", *Russian Math. Surveys* **67**:5 (2012), 785–890.

[32] S. G. Bobkov and M. Ledoux, *One-dimensional empirical measures, order statistics and Kantorovich transport distances*, Mem. Amer. Math. Soc. (to appear); Preprint, 2014, 140 pp., http://math.umn.edu/~bobko001/preprints/2014_BL_Order.statistics.13.pdf.

[33] Л. В. Канторович, Г. П. Акилов, *Функциональный анализ*, 3-е изд., Наука, М. 1984, 752 с.; English transl. of 2nd ed., L. V. Kantorovich and G. P. Akilov, *Functional analysis*, 2nd ed., Pergamon Press, Oxford–Elmsford, NY 1982, xiv+589 pp.

[34] R. M. Dudley, *Real analysis and probability*, The Wadsworth & Brooks/Cole Math. Ser., The Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA 1989, xii+436 pp.

[35] E. Rio, "Upper bounds for minimal distances in the central limit theorem", *Ann. Inst. Henri Poincaré Probab. Stat.* **45**:3 (2009), 802–817.

[36] M. Talagrand, "Transportation cost for Gaussian and other product measures", *Geom. Funct. Anal.* **6**:3 (1996), 587–600.

[37] S. G. Bobkov and F. Götze, "Exponential integrability and transportation cost related to logarithmic Sobolev inequalities", *J. Funct. Anal.* **163**:1 (1999), 1–28.

[38] S. G. Bobkov, N. Gozlan, C. Roberto, and P.-M. Samson, "Bounds on the deficit in the logarithmic Sobolev inequality", *J. Funct. Anal.* **267**:11 (2014), 4110–4138.

[39] A. E. Ingham, "A note on Fourier transforms", *J. London Math. Soc.* **9**:1 (1934), 29–32.

[40] R. N. Bhattacharya and R. Ranga Rao, *Normal approximation and asymptotic expansions*, Wiley Ser. Probab. Stat., John Wiley & Sons, Inc., New York–London–Sydney 1976, xiv+274 pp.; updated republication of the 1986 reprint, Classics Appl. Math., vol. 64, SIAM, Philadelphia, PA 2010, xxii+316 pp.

[41] W. Gautschi, "On inverses of Vandermonde and confluent Vandermonde matrices", *Numer. Math.* **4** (1962), 117–123.

[42] С. Г. Бобков, "К одной теореме В. Н. Судакова о типичных распределениях", *Вероятность и статистика*. 15, Зап. науч. сем. ПОМИ, **368**, ПОМИ, СПб. 2009, с. 59–74; English transl., S. G. Bobkov, "On a theorem of V. N. Sudakov on typical distributions", *J. Math. Sci. (N. Y.)* **167**:4 (2010), 464–473.

[43] I. Pinelis, *On the nonuniform Berry–Esseen bound*, 2013, 20 pp., arXiv: 1301.2828.

[44] I. Pinelis, *More on the nonuniform Berry–Esseen bound*, 2013, 8 pp., arXiv: 1302.0516.

[45] H. Prawitz, "Limits for a distribution, if the characteristic function is given in a finite domain", *Skand. Aktuarietidskr.* **1972** (1973), 138–154.

[46] H. Bohman, "To compute the distribution function when the characteristic function is known", *Skand. Aktuarietidskr.* **1963** (1964), 41–46.

[47] J. D. Vaaler, "Some extremal functions in Fourier analysis", *Bull. Amer. Math. Soc. (N. S.)* **12**:2 (1985), 183–216.

[48] В. К. Мацкявичюс, "О нижней оценке скорости сходимости в центральной предельной теореме", *Теория вероятн. и ее примен.* **28**:3 (1983), 565–569; English transl., V. K. Matskyavichyus, "A lower bound for the convergence rate in the central limit theorem", *Theory Probab. Appl.* **28**:3 (1984), 596–601.

[49] И. Г. Шевцова, "Об абсолютных константах в неравенстве типа Берри–Эссеена", *Докл. РАН* **456**:6 (2014), 650–654; English transl., I. G. Shevtsova, "On the absolute constants in the Berry–Esseen-type inequalities", *Dokl. Math.* **89**:3 (2014), 378–381.

[50] Л. Д. Мешалкин, Б. А. Рогозин, "Оценка расстояния между функциями распределения по близости их характеристических функций и ее применение в центральной предельной теореме", *Предельные теоремы теории вероятностей*, Изд-во АН УзССР, Ташкент 1963, с. 40–55. [L. D. Meshalkin and B. A. Rogozin, "Estimate for the distance between distribution functions on the basis of the proximity of their characteristic functions and its applications to the central limit theorem", *Limit theorems in probability theory*, Publishing House of the Academy of Sciences of Uzbekistan Soviet Republic, Tashkent 1963, pp. 40–55.]

[51] С. В. Нагаев, "Некоторые предельные теоремы для больших уклонений", *Теория вероятн. и ее примен.* **10**:2 (1965), 231–254; English transl.,

S. V. Nagaev, “Some limit theorems for large deviations”, *Theory Probab. Appl.* **10**:2 (1965), 214–235.

[52] A. Бикялис, “Оценки остаточного члена в центральной предельной теореме”, *Литов. матем. сб.* **6**:3 (1966), 323–346. [A. Bikelis, “Estimates of the remainder term in the central limit theorem”, *Litovskii Mat. Sb.* **6**:3 (1966), 323–346.]

[53] L. Paditz, “Abschätzungen der Konvergenzgeschwindigkeit zur Normalverteilung unter Voraussetzung einseitiger Momente”, *Math. Nachr.* **82** (1978), 131–156.

[54] L. Paditz, “On the analytical structure of the constant in the nonuniform version of the Esseen inequality”, *Statistics* **20**:3 (1989), 453–464.

[55] R. Michel, “On the constant in the nonuniform version of the Berry–Esseen theorem”, *Z. Wahrscheinlichkeitstheor. verw. Geb.* **55**:1 (1981), 109–117.

[56] Ю. С. Нефедова, И. Г. Шевцова, “О точности нормальной аппроксимации для распределений пуассоновских случайных сумм”, *Информ. и ее примен.* **5**:1 (2011), 39–45. [Yu. S. Nefedova and I. G. Shevtsova, “Accuracy of normal approximation to distributions of Poisson random sums”, *Informatika i Primenen.* **5**:1 (2011), 39–45.]

[57] Ю. С. Нефедова, И. Г. Шевцова, “О неравномерных оценках скорости сходимости в центральной предельной теореме”, *Теория вероятн. и ее примен.* **57**:1 (2012), 62–97; English transl., Yu. S. Nefedova and I. G. Shevtsova, “On nonuniform convergence rate estimates in the central limit theorem”, *Theory Probab. Appl.* **57**:1 (2013), 28–59.

[58] E. Rio, “Asymptotic constants for minimal distance in the central limit theorem”, *Electron. Commun. Probab.* **16** (2011), 96–103.

[59] S. G. Bobkov, “Entropic approach to E. Rio's central limit theorem for W_2 transport distance”, *Statist. Probab. Lett.* **83**:7 (2013), 1644–1648.

[60] S. G. Bobkov, “Berry–Esseen bounds and Edgeworth expansions in the central limit theorem for transport distances”, *Probab. Theory Related Fields* (to appear); Preprint, 2016, 29 pp., http://math.umn.edu/~bobko001/preprints/CLT.for.W_p.6_Revision.pdf.

[61] В. В. Сенатов, *Центральная предельная теорема. Точность аппроксимации и асимптотические разложения*, Книжный дом ЛИБРОКОМ, М. 2009, 352 с. [V. V. Senatov, *The central limit theorem. Accuracy of approximation and asymptotic expansions*, Knizhnyi Dom LIBROKOM, Moscow 2009, 352 pp.]

Sergei G. Bobkov
School of Mathematics, University of Minnesota,
Minneapolis, MN, USA
E-mail: bobkov@math.umn.edu

Received 30/DEC/15
Translated by N. BERESTOVA