Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 126 (2016) 3865–3887

stochastic processes and their applications

www.elsevier.com/locate/spa

Regularized distributions and entropic stability of Cramer's characterization of the normal law

S.G. Bobkov[a], G.P. Chistyakov[b], F. Götze[b,*]

[a] *School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455, USA*
[b] *Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany*

Received 2 January 2015; accepted 5 July 2015
Available online 29 April 2016

Abstract

For regularized distributions we establish stability of the characterization of the normal law in Cramer's theorem with respect to the total variation norm and the entropic distance. As part of the argument, Sapogov-type theorems are refined for random variables with finite second moment.

MSC: primary 60E

Keywords: Cramer's theorem; Normal characterization; Stability problems

1. Introduction

Let X and Y be independent random variables. A theorem of Cramer [8] indicates that, if the sum $X + Y$ has a normal distribution, then both X and Y are normal. P. Lévy established stability of this characterization property with respect to the Lévy distance, which is formulated as follows. Given $\varepsilon > 0$ and distribution functions F, G,

$$L(F * G, \Phi) < \varepsilon \Rightarrow L(F, \Phi_{a_1,\sigma_1}) < \delta_\varepsilon, L(G, \Phi_{a_2,\sigma_2}) < \delta_\varepsilon,$$

* Corresponding author.
E-mail addresses: bobkov@math.umn.edu (S.G. Bobkov), chistyak@math.uni-bielefeld.de (G.P. Chistyakov), goetze@math.uni-bielefeld.de (F. Götze).

http://dx.doi.org/10.1016/j.spa.2016.04.010
0304-4149/

for some $a_1, a_2 \in \mathbf{R}$ and $\sigma_1, \sigma_2 > 0$, where δ_ε only depends on ε, and in such a way that $\delta_\varepsilon \to 0$ as $\varepsilon \to 0$. Here $\Phi_{a,\sigma}$ stands for the distribution function of the normal law $N(a, \sigma^2)$ with mean a and standard deviation σ, i.e., with density

$$\varphi_{a,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-a)^2/2\sigma^2}, \quad x \in \mathbf{R},$$

and we omit indices in the standard case $a = 0, \sigma = 1$. As usual, $F * G$ denotes the convolution of the corresponding distributions.

The problem of quantitative versions of this stability property of the normal law has been intensively studied in many papers, starting with results by Sapogov [18,19] and ending with results by Chistyakov and Golinskii [7], who found the correct asymptotic of the best possible error function $\varepsilon \to \delta_\varepsilon$ for the Lévy distance. See also [13–16,6,21–24].

As for stronger metrics, not much is known up to now. According to McKean ([17], cf. also [5] for some related aspects of the problem), it was Kac who raised the question about the stability in Cramer's theorem with respect to the entropic distance to normality. Let us recall that, if a random variable X with finite second moment has a density $p(x)$, its entropy

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

is well-defined and is bounded from above by the entropy of the normal random variable Z, having the same variance $\sigma^2 = \mathrm{Var}(Z) = \mathrm{Var}(X)$. The entropic distance to the normal is given by the formula

$$D(X) = h(Z) - h(X) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{\varphi_{a,\sigma}(x)}\, dx,$$

where in the last formula it is assumed that $a = \mathbf{E}Z = \mathbf{E}X$. It represents the Kullback–Leibler distance from the distribution F of X to the family of all normal laws on the line.

In general, $0 \le D(X) \le \infty$, and an infinite value is possible. This quantity is affine invariant, and so it does not depend on the mean and variance of X. It is stronger than the total variation distance $\|F - \Phi_{a,\sigma}\|_{\mathrm{TV}}$, as may be seen from the Pinsker inequality

$$D(X) \ge \frac{1}{2}\, \|F - \Phi_{a,\sigma}\|_{\mathrm{TV}}^2.$$

Thus, Kac's question is whether one can bound the entropic distance $D(X+Y)$ from below in terms of $D(X)$ and $D(Y)$ for independent random variables, i.e., to have an inequality

$$D(X+Y) \ge \alpha(D(X), D(Y))$$

with some non-negative function α, such that $\alpha(t,s) > 0$ for $t, s > 0$. If so, Cramer's theorem would be an immediate consequence of this. Note that the reverse inequality does exist, and in case $\mathrm{Var}(X+Y) = 1$ we have

$$D(X+Y) \le \mathrm{Var}(X) D(X) + \mathrm{Var}(Y) D(Y),$$

which is due to the general entropy power inequality, cf. [9].

It turned out that Kac's question has a negative solution. More precisely, for any $\varepsilon > 0$, one can construct independent random variables X and Y with absolutely continuous symmetric distributions F, G, and with $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$, such that

(a) $D(X+Y) < \varepsilon$;
(b) $\|F - \Phi_{a,\sigma}\|_{\rm TV} > c$ and $\|G - \Phi_{a,\sigma}\|_{\rm TV} > c$, for all $a \in \mathbf{R}$ and $\sigma > 0$,

where $c > 0$ is an absolute constant, see [1]. In particular, $D(X)$ and $D(Y)$ are bounded away from zero. Moreover, refined analytic tools show that the random variables may be chosen to be identically distributed, i.e., (a)–(b) hold with $F = G$, see [2].

Nevertheless, Kac's problem remains to be of interest for subclasses of probability measures obtained by convolution with a "smooth" distribution. The main purpose of this note is to give an affirmative solution to the problem in the (rather typical) situation, when independent Gaussian noise is added to the given random variables. That is, for a small parameter $\sigma > 0$, we consider the regularized random variables $X_\sigma = X + \sigma Z_1$, $Y_\sigma = Y + \sigma Z_2$, where Z_1 and Z_2 denote independent standard normal random variables, which are independent of X, Y. Note that the functionals $D(X_\sigma)$, $D(Y_\sigma)$, $D(X_\sigma + Y_\sigma)$ are still translation invariant. As a main result, we prove:

Theorem 1.1. *Let X, Y be independent random variables with* $\mathrm{Var}(X+Y) = 1$. *Given* $0 < \sigma \le 1$, *the regularized random variables X_σ and Y_σ satisfy*

$$D(X_\sigma + Y_\sigma) \ge \exp\Big\{-\frac{c\,\log^7(2+1/D)}{D^2}\Big\},$$

where $c > 0$ is an absolute constant, and

$$D = \sigma^2 \left(\mathrm{Var}(X_\sigma)\, D(X_\sigma) + \mathrm{Var}(Y_\sigma)\, D(Y_\sigma)\right).$$

Thus, if $D(X_\sigma + Y_\sigma)$ is small, the entropic distances $D(X_\sigma)$ and $D(Y_\sigma)$ have to be small, as well. In particular, Cramer's theorem is a consequence of this statement. However, it is not clear whether the above lower bound is optimal with respect to the couple $(D(X_\sigma), D(Y_\sigma))$, and perhaps the logarithmic term in the exponent may be removed. As we will see, a certain improvement of the bound can be achieved, when X and Y have equal variances.

Beyond the realm of results around P. Lévy's theorem, recently there has been renewed interest in other related stability problems in different areas of Analysis and Geometry. One can mention, for example, the problems of sharpness of the Brunn–Minkowski and Sobolev-type inequalities (cf. [10,11,20,4]).

We start with the description and refinement of Sapogov-type theorems about the normal approximation in Kolmogorov distance (Sections 2 and 3) and then turn to analogous results for the Lévy distance (Section 4). A version of Theorem 1.1 for the total variation distance is given in Section 5. Sections 6 and 7 deal with the problem of bounding the tail function $\mathbf{E}X^2\, 1_{\{|X| \ge T\}}$ in terms of the entropic distances $D(X)$ and $D(X+Y)$, which is an essential part of Kac's problem. A first application, namely, to a variant of Chistyakov–Golinskii's theorem, is discussed in Section 8. In Section 9, we develop several estimates connecting the entropic distance $D(X)$ and the uniform deviation of the density p from the corresponding normal density. In Section 10 an improved variant of Theorem 1.1 is derived in the case, where X and Y have equal variances. The general case is treated in Section 11. Finally, some relations between different distances in the space of probability distributions on the line are postponed to Appendix (without proofs), and we refer to the extended version [3] for more details.

2. Sapogov-type theorems for Kolmogorov distance

Throughout the paper we consider the following classical metrics in the space of probability distributions on the real line:

(1) The Kolmogorov or L^∞-distance $\|F - G\| = \sup_x |F(x) - G(x)|$;
(2) The Lévy distance

$$L(F, G) = \min\big\{h \ge 0 : G(x - h) - h \le F(x) \le G(x + h) + h,\ \forall x \in \mathbf{R}\big\};$$

(3) The Kantorovich or L^1-distance $W_1(F, G) = \int_{-\infty}^{\infty} |F(x) - G(x)|\, dx$;
(4) The total variation distance

$$\|F - G\|_{\mathrm{TV}} = \sup \sum |(F(x_k) - G(x_k)) - (F(y_k) - G(y_k))|,$$

where the sup is taken over all finite collections of points $y_1 < x_1 < \cdots < y_n < x_n$.

In these relations, F and G are arbitrary distribution functions. Note that the quantity $W_1(F, G)$ is finite, as long as both F and G have a finite first absolute moment.

In the sequel, $\Phi_{a,v}$ or $N(a, v^2)$ denote the normal distribution (function) with parameters (a, v^2), $a \in \mathbf{R} v > 0$. If $a = 0$, we write Φ_v, and write Φ in the standard case $a = 0$, $v = 1$.

Now, let X and Y be independent random variables with distribution functions F and G. Then the convolution $F * G$ represents the distribution of the sum $X + Y$. If both random variables have mean zero and unit variances, Sapogov's main stability result reads as follows:

Theorem 2.1. *Let* $\mathbf{E}X = \mathbf{E}Y = 0$ *and* $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$. *If* $\|F * G - \Phi * \Phi\| \le \varepsilon < 1$, *then with some absolute constant* C

$$\|F - \Phi\| \le \frac{C}{\sqrt{\log \frac{1}{\varepsilon}}} \quad \text{and} \quad \|G - \Phi\| \le \frac{C}{\sqrt{\log \frac{1}{\varepsilon}}}.$$

In the general case (that is, when there are no finite moments), the conclusion is somewhat weaker. Namely, with $\varepsilon \in (0, 1)$, we associate

$$a_1 = \int_{-N}^{N} x\, dF(x), \qquad \sigma_1^2 = \int_{-N}^{N} x^2\, dF(x) - a_1^2 \quad (\sigma_1 \ge 0),$$

and similarly (a_2, σ_2^2) for the distribution function G, where $N = N(\varepsilon) = 1 + \sqrt{2\log(1/\varepsilon)}$. In the sequel, we also use the function $m(\sigma, \varepsilon) = \min\{\frac{1}{\sqrt{\sigma}}, \log\log \frac{e^e}{\varepsilon}\}$, $\sigma > 0$, $0 < \varepsilon \le 1$.

Theorem 2.2. *Assume* $\|F * G - \Phi\| \le \varepsilon < 1$. *If* F *has median zero, and* $\sigma_1, \sigma_2 > 0$, *then with some absolute constant* C

$$\|F - \Phi_{a_1,\sigma_1}\| \le \frac{C}{\sigma_1 \sqrt{\log \frac{1}{\varepsilon}}}\, m(\sigma_1, \varepsilon),$$

and similarly for G.

Originally, Sapogov derived a weaker bound in [18] with worse behavior with respect to both σ_1 and ε. In [19] he gave an improvement, $\|F - \Phi_{a_1,\sigma_1}\| \le \frac{C}{\sigma_1^3 \sqrt{\log(1/\varepsilon)}}$ with a correct asymptotic of the right-hand side with respect to ε, cf. also [15]. The correctness of the asymptotic with respect to ε was studied in [16], cf. also [6]. In 1976 Senatov [21], using the ridge property of characteristic functions, improved the factor σ_1^3 to $\sigma_1^{3/2}$, i.e.,

$$\|F - \Phi_{a_1,\sigma_1}\| \le \frac{C}{\sigma_1^{3/2} \sqrt{\log \frac{1}{\varepsilon}}}. \tag{2.1}$$

He also emphasized that the presence of σ_1 in the bound is essential. A further improvement of the power of σ_1 is due to Shiganov [22,23]. Moreover, at the expense of an additional ε-dependent factor, one can replace $\sigma_1^{3/2}$ with σ_1. As shown in [7], see Remark on p. 2861,

$$\|F - \Phi_{a_1,\sigma_1}\| \le \frac{C \log\log \frac{e^e}{\varepsilon}}{\sigma_1 \sqrt{\log \frac{1}{\varepsilon}}}. \tag{2.2}$$

Therefore, Theorem 2.2 is just the combination of the two results, (2.1) and (2.2).

Let us emphasize that all proofs of these theorems use the methods of the Complex Analysis. Moreover, up to now there is no "Real Analysis" proof of the Cramér theorem and of its extensions in the form of Sapogov-type results. This, however, does not concern the case of identically distributed summands, cf. [2].

We will discuss the bounds in the Lévy distance in the next sections.

The assumption about the median in Theorem 2.2 may be weakened to the condition that the medians of X and Y, $m(X)$ and $m(Y)$, are bounded in absolute value by a constant. For example, if $\mathbf{E}X = \mathbf{E}Y = 0$ and $\mathrm{Var}(X+Y) = 1$, and if, for definiteness, $\mathrm{Var}(X) \le 1/2$, then, by Chebyshev's inequality, $|m(X)| \le 1$, while $|m(Y)|$ will be bounded by an absolute constant, when ε is small enough, due to the main hypothesis $\|F * G - \Phi\| \le \varepsilon$.

Moreover, if the variances of X and Y are bounded away from zero, the statement of Theorem 2.2 holds with $a_1 = 0$, and the factor σ_1 can be replaced with the standard deviation of X. In the next section, we recall some standard arguments in order to justify this conclusion and give a more general version of Theorem 2.2 involving variances:

Theorem 2.3. *Let* $\mathbf{E}X = \mathbf{E}Y = 0$, $\mathrm{Var}(X+Y) = 1$. *If* $\|F * G - \Phi\| \le \varepsilon < 1$, *then with some absolute constant* C

$$\|F - \Phi_{v_1}\| \le \frac{C m(v_1,\varepsilon)}{v_1 \sqrt{\log \frac{1}{\varepsilon}}} \quad \text{and} \quad \|G - \Phi_{v_2}\| \le \frac{C m(v_2,\varepsilon)}{v_2 \sqrt{\log \frac{1}{\varepsilon}}},$$

where $v_1^2 = \mathrm{Var}(X)$, $v_2^2 = \mathrm{Var}(Y)$ $(v_1, v_2 > 0)$.

Under the stated assumptions, Theorem 2.3 is stronger than Theorem 2.2, since $v_1 \ge \sigma_1$. Another advantage of this formulation is that v_1 does not depend on ε, while σ_1 does.

3. Proof of Theorem 2.3

Let X and Y be independent random variables with distribution functions F and G, respectively, with $\mathbf{E}X = \mathbf{E}Y = 0$ and $\mathrm{Var}(X+Y) = 1$. We assume that $\|F * G - \Phi\| \le \varepsilon < 1$, and keep the same notations as in Section 2. Recall that $N = N(\varepsilon) = 1 + \sqrt{2\log(1/\varepsilon)}$.

The proof of Theorem 2.3 is entirely based on Theorem 2.2. We will need:

Lemma 3.1. *With some absolute constant* C *we have* $0 \le 1 - (\sigma_1^2 + \sigma_2^2) \le C N^2 \sqrt{\varepsilon}$.

A similar assertion, $|\sigma_1^2 + \sigma_2^2 - 1| \le C N^2 \varepsilon$, is known under the assumption that F has a median at zero (without moment assumptions). For the proof of Lemma 3.1, we use arguments from [18,21], cf. Lemma 1. It will be convenient to divide the proof into several steps.

Lemma 3.2. *Let* $\varepsilon \le \varepsilon_0 = \frac{1}{4} - \Phi(-1) = 0.0913\ldots$ *Then* $|m(X)| \le 2$ *and* $|m(Y)| \le 2$.

Indeed, let $\mathrm{Var}(X) \le 1/2$. Then $|m(X)| \le 1$, by Chebyshev's inequality. Hence,

$$\frac{1}{4} \le \mathbf{P}\{X \le 1,\ Y \le m(Y)\} \le \mathbf{P}\{X + Y \le m(Y) + 1\} \le \Phi(m(Y) + 1) + \varepsilon,$$

which for $\varepsilon \le \frac{1}{4}$ implies that $m(Y) + 1 \ge \Phi^{-1}(\frac{1}{4} - \varepsilon)$. In particular, $m(Y) \ge -2$, if $\varepsilon \le \varepsilon_0$. Similarly, $m(Y) \le 2$. □

To continue, introduce truncated random variables at level N. Put $X^* = X$ in case $|X| \le N$, $X^* = 0$ in case $|X| > N$, and similarly Y^* for Y. Note that $\mathbf{E}X^* = a_1$, $\mathrm{Var}(X^*) = \sigma_1^2$, $\mathbf{E}Y^* = a_2$, $\mathrm{Var}(Y^*) = \sigma_2^2$. By the construction, $\sigma_1 \le v_1$ and $\sigma_2 \le v_2$. In particular, $\sigma_1^2 + \sigma_2^2 \le v_1^2 + v_2^2 = 1$. Let F^*, G^* denote the distribution functions of X^*, Y^*, respectively.

Lemma 3.3. *With some absolute constant C we have $\|F^* - F\| \le C\sqrt{\varepsilon}$, $\|G^* - G\| \le C\sqrt{\varepsilon}$, $\|F^* * G^* - \Phi\| \le C\sqrt{\varepsilon}$.*

Proof. One may assume that $N = N(\varepsilon)$ is a point of continuity of both F and G. Since the Kolmogorov distance is bounded by 1, one may also assume that ε is sufficiently small, e.g., $\varepsilon < \min\{\varepsilon_0, \varepsilon_1\}$, where $\varepsilon_1 = \exp\{-1/(3 - 2\sqrt{2})\}$. In this case $(N-2)^2 > (N-1)^2/2$, so

$$\Phi(-(N-2)) = 1 - \Phi(N-2) \le \frac{1}{2}\, e^{-(N-2)^2/2} \le \frac{1}{2}\, e^{-(N-1)^2/4} = \frac{\sqrt{\varepsilon}}{2}.$$

By Lemma 3.2 and the basic assumption on the convolution $F * G$,

$$\frac{1}{2}\,\mathbf{P}\{Y \le -N\} \le \mathbf{P}\{X \le 2,\ Y \le -N\}$$
$$\le \mathbf{P}\{X + Y \le -(N-2)\} = (F * G)(-(N-2)) \le \Phi(-(N-2)) + \varepsilon.$$

So, $G(-N) \le 2\Phi(-(N-2)) + 2\varepsilon \le 3\sqrt{\varepsilon}$. Analogously, $1 - G(N) \le 3\sqrt{\varepsilon}$. Thus, $\int_{\{|x| \ge N\}} dG(x) \le 6\sqrt{\varepsilon}$ as well as $\int_{\{|x| \ge N\}} dF(x) \le 6\sqrt{\varepsilon}$.

In particular, for $x < -N$, we have $|F^*(x) - F(x)| = F(x) \le 6\sqrt{\varepsilon}$, and similarly for $x > N$. If $-N < x < 0$, then $F^*(x) = F(x) - F(-N)$, and if $0 < x < N$, we have $F^*(x) = F(x) + (1 - F(N))$. In both cases, $|F^*(x) - F(x)| \le 6\sqrt{\varepsilon}$. Therefore, $\|F^* - F\| \le 6\sqrt{\varepsilon}$. Similarly, $\|G^* - G\| \le 6\sqrt{\varepsilon}$. From this, by the triangle inequality,

$$\|F^* * G^* - F * G\| \le \|F^* * G^* - F^* * G\| + \|F^* * G - F * G\|$$
$$\le \|F^* - F\| + \|G^* - G\| \le 12\sqrt{\varepsilon}.$$

Finally, $\|F^* * G^* - \Phi\| \le \|F^* * G^* - F * G\| + \|F * G - \Phi\| \le 12\sqrt{\varepsilon} + \varepsilon \le 13\sqrt{\varepsilon}$. □

Proof of Lemma 3.1. Since $|X^* + Y^*| \le 2N$ and $a_1 + a_2 = \mathbf{E}\,(X^* + Y^*) = \int x\, dF^* * G^*(x)$, we have, integrating by parts,

$$a_1 + a_2 = \int_{-2N}^{2N} x\, d((F^* * G^*)(x) - \Phi(x))$$
$$= x\,((F^* * G^*)(x) - \Phi(x))\Big|_{x=-2N}^{x=2N} - \int_{-2N}^{2N} ((F^* * G^*)(x) - \Phi(x))\, dx.$$

Hence, $|a_1 + a_2| \le 8N \, \|F^* * G^* - \Phi\|$, which, by Lemma 3.3, is bounded by $CN\sqrt{\varepsilon}$. Similarly,

$$\begin{aligned}\mathbf{E}\,(X^* + Y^*)^2 - 1 &= \int_{-2N}^{2N} x^2 \, d((F^* * G^*)(x) - \Phi(x)) - \int_{\{|x|>2N\}} x^2 \, d\Phi(x) \\ &= x^2 \, ((F^* * G^*)(x) - \Phi(x))\Big|_{x=-2N}^{x=2N} \\ &\quad - 2\int_{-2N}^{2N} x \, ((F^* * G^*)(x) - \Phi(x)) \, dx - \int_{\{|x|>2N\}} x^2 \, d\Phi(x).\end{aligned}$$

Hence,

$$\Big|\mathbf{E}\,(X^* + Y^*)^2 - 1\Big| \le 24\, N^2 \, \|F^* * G^* - \Phi\| + 2\int_{2N}^{\infty} x^2 \, d\Phi(x).$$

The last integral asymptotically behaves like $2N\varphi(2N) < Ne^{-2(N-1)^2} = N\varepsilon^4$. Therefore, $\big|\mathbf{E}\,(X^* + Y^*)^2 - 1\big|$ is bounded by $CN^2\sqrt{\varepsilon}$. Finally, writing $\sigma_1^2 + \sigma_2^2 = \mathbf{E}\,(X^* + Y^*)^2 - (a_1 + a_2)^2$, we get that

$$\Big|\sigma_1^2 + \sigma_2^2 - 1\Big| \le \Big|\mathbf{E}\,(X^* + Y^*)^2 - 1\Big| + (a_1 + a_2)^2 \le CN^2\sqrt{\varepsilon}$$

with some absolute constant C. Lemma 3.1 follows. □

Proof of Theorem 2.3. First note that, given $a > 0, \sigma > 0$, and $x \in \mathbf{R}$, the function $\psi(x) = \Phi_{0,\sigma}(x) - \Phi_{a,\sigma}(x) = \Phi\big(\frac{x}{\sigma}\big) - \Phi\big(\frac{x-a}{\sigma}\big)$ is vanishing at infinity, has a unique extreme point $x_0 = \frac{a}{2}$, and $\psi(x_0) = \int_{-a/2\sigma}^{a/2\sigma} \varphi(y) \, dy \le \frac{a}{\sigma\sqrt{2\pi}}$. Hence, including the case $a \le 0$, as well, we get

$$\|\Phi_{a,\sigma} - \Phi_{0,\sigma}\| \le \frac{|a|}{\sigma\sqrt{2\pi}}.$$

We apply this estimate for $a = a_1$ and $\sigma = \sigma_1$. Since $\mathbf{E}X = 0$ and $\mathrm{Var}(X + Y) = 1$, by Cauchy's and Chebyshev's inequalities,

$$|a_1| = \big|\mathbf{E}\, X \, 1_{\{|X| \ge N\}}\big| \le \mathbf{P}\{|X| \ge N\}^{1/2} \le \frac{1}{N} < \frac{1}{\sqrt{\log \frac{e}{\varepsilon}}}.$$

Hence, $\|\Phi_{a_1,\sigma_1} - \Phi_{0,\sigma_1}\| \le \frac{|a_1|}{\sigma_1\sqrt{2\pi}} \le \frac{C}{\sigma_1\sqrt{\log \frac{1}{\varepsilon}}}$. A similar inequality also holds for (a_2, σ_2).

Now, define the non-negative numbers $u_1 = v_1 - \sigma_1$, $u_2 = v_2 - \sigma_2$. By Lemma 3.1,

$$\begin{aligned}CN^2\sqrt{\varepsilon} \ge 1 - (\sigma_1^2 + \sigma_2^2) &= 1 - \Big((v_1 - u_1)^2 + (v_2 - u_2)^2\Big) \\ &= u_1\,(2v_1 - u_1) + u_2\,(2v_2 - u_2) \ge u_1 v_1 + u_2 v_2.\end{aligned}$$

Hence, $u_1 \le \frac{CN^2\sqrt{\varepsilon}}{v_1}$, $u_2 \le \frac{CN^2\sqrt{\varepsilon}}{v_2}$. These relations can be used to estimate $\Delta = \|\Phi_{0,v_1} - \Phi_{0,\sigma_1}\|$. Given two parameters $\alpha > \beta > 0$, consider the function of the form $\psi(x) = \Phi(\alpha x) - \Phi(\beta x)$. In case $x > 0$, by the mean value theorem, for some $x_0 \in (\beta x, \alpha x)$,

$$\psi(x) = (\alpha - \beta)\, x\varphi(x_0) < (\alpha - \beta)\, x\varphi(\beta x).$$

Here, the right-hand side is maximized for $x = \frac{1}{\beta}$, which gives $\psi(x) < \frac{1}{\sqrt{2\pi e}} \frac{\alpha-\beta}{\beta}$. A similar bound also holds for $x < 0$. Using this bound with $\alpha = 1/\sigma_1$ ($\sigma_1 > 0$), $\beta = 1/v_1$, we obtain

$$\Delta \leq \frac{1}{\sqrt{2\pi e}} v_1 \left(\frac{1}{\sigma_1} - \frac{1}{v_1} \right) = \frac{1}{\sqrt{2\pi e}} \frac{u_1}{\sigma_1} \leq \frac{CN^2\sqrt{\varepsilon}}{\sigma_1 v_1} \leq \frac{CN^2\sqrt{\varepsilon}}{\sigma_1^2}.$$

Thus, applying Theorem 2.2, we get with some universal constant $C > 1$ that

$$\begin{aligned} \|F - \Phi_{0,v_1}\| &\leq \|F - \Phi_{a_1,\sigma_1}\| + \|\Phi_{a_1,\sigma_1} - \Phi_{0,\sigma_1}\| + \|\Phi_{0,\sigma_1} - \Phi_{0,v_1}\| \\ &\leq \frac{C}{\sigma_1\sqrt{\log\frac{1}{\varepsilon}}} m(\sigma_1,\varepsilon) + \frac{C}{\sigma_1\sqrt{\log\frac{1}{\varepsilon}}} + \frac{CN^2\sqrt{\varepsilon}}{\sigma_1^2} \\ &\leq \frac{2C}{\sigma_1\sqrt{\log\frac{1}{\varepsilon}}} m(\sigma_1,\varepsilon) + \frac{CN^2\sqrt{\varepsilon}}{\sigma_1^2}. \end{aligned} \tag{3.1}$$

The obtained estimate remains valid when $\sigma_1 = 0$, as well. On the other hand, $\sigma_1 = v_1 - u_1 \geq v_1 - \frac{CN^2\sqrt{\varepsilon}}{v_1} \geq \frac{1}{2} v_1$ where the last inequality is fulfilled for the range $v_1 \geq v(\varepsilon) = \sqrt{C}\, N\, (4\varepsilon)^{1/4}$. Hence, from (3.1) and using $m(\sigma_1,\varepsilon) \leq 2m(v_1,\varepsilon)$, for this range

$$\|F - \Phi_{0,v_1}\| \leq \frac{8Cm(v_1,\varepsilon)}{v_1\sqrt{\log\frac{1}{\varepsilon}}} + \frac{4CN^2\sqrt{\varepsilon}}{v_1^2}.$$

Here, since $m(v_1,\varepsilon) \geq 1$, the first term on the right-hand side majorizes the second one, if $v_1 \geq \tilde{v}(\varepsilon) = N^2\sqrt{\varepsilon\log\frac{1}{\varepsilon}}$. Therefore, when $v_1 \geq w(\varepsilon) = \max\{v(\varepsilon), \tilde{v}(\varepsilon)\}$, with some absolute constant C' we have

$$\|F - \Phi_{0,v_1}\| \leq \frac{C' m(v_1,\varepsilon)}{v_1\sqrt{\log\frac{1}{\varepsilon}}}.$$

Thus, we arrive at the desired inequality for the range $v_1 \geq w(\varepsilon)$. But the function w behaves almost polynomially near zero and admits, for example, a bound of the form $w(\varepsilon) \leq \sqrt{C''}\,\varepsilon^{1/6}$, $0 < \varepsilon < \varepsilon_0$, with some universal $\varepsilon_0 \in (0,1)$, $C'' > 1$. So, when $v_1 \leq w(\varepsilon)$, $0 < \varepsilon < \varepsilon_0$, we have

$$\frac{1}{v_1\sqrt{\log\frac{1}{\varepsilon}}} \geq \frac{1}{w(\varepsilon)\sqrt{\log\frac{1}{\varepsilon}}} \geq \frac{1}{\varepsilon^{1/6}\sqrt{C''\log\frac{1}{\varepsilon}}}.$$

Here, the last expression is greater than 1, as long as ε is sufficiently small, say, for all $0 < \varepsilon < \varepsilon_1$, where ε_1 is determined by (C'', ε_0). Hence, for all such ε, we have a better bound $\|F - \Phi_{0,v_1}\| \leq \frac{C}{v_1\sqrt{\log\frac{1}{\varepsilon}}}$. It remains to increase the constant C' in order to involve the remaining values of ε. A similar conclusion is true for G. □

4. Stability in Cramer's theorem for the Lévy distance

Let X and Y be independent random variables with distribution functions F and G. It turns out that in the bound of Theorem 2.2, the parameter σ_1 can be completely removed, if we consider the stability problem for the Lévy distance. More precisely, the following theorem was established in [7].

Theorem 4.1. *Assume that* $\|F*G-\Phi\| \le \varepsilon < 1$. *If* F *has median zero, then with some absolute constant* C

$$L(F, \Phi_{a_1,\sigma_1}) \le C\, \frac{\left(\log\log\frac{4}{\varepsilon}\right)^2}{\sqrt{\log\frac{1}{\varepsilon}}}.$$

Recall that $a_1 = \int_{-N}^{N} x\, dF(x)$, $\sigma_1^2 = \int_{-N}^{N} x^2\, dF(x) - a_1^2 (\sigma_1 \ge 0)$, and similarly (a_2, σ_2^2) for G, where $N = 1+\sqrt{2\log(1/\varepsilon)}$. As we have already discussed, the assumption about the median may be relaxed to the condition that the median is bounded (by a universal constant).

The first quantitative stability result for the Lévy distance, namely,

$$L(F, \Phi_{a_1,\sigma_1}) \le C\, \log^{-1/8}(1/\varepsilon),$$

was obtained in 1968 by Zolotarev [24], who applied his famous Berry–Esseen-type bound [25]. The power 1/8 was later improved to 1/4 by Senatov [21] and even more by Shiganov [22,23]. The stated asymptotic in Theorem 4.1 is unimprovable, which was also shown in [7].

Note that in the assumption of Theorem 4.1, the Kolmogorov distance can be replaced with the Lévy distance $L(F, \Phi)$ in view of the general relations $L(F, \Phi) \le \|F * G - \Phi\| \le (1 + M)\, L(F, \Phi)$ with $M = \|\Phi\|_{\mathrm{Lip}} = \frac{1}{\sqrt{2\pi}}$. However, in the conclusion such replacement cannot be done at the expense of a universal constant, since we only have

$$\|F - \Phi_{a_1,\sigma_1}\| \le (1 + M)\, L(F, \Phi_{a_1,\sigma_1}), \qquad M = \|\Phi_{a_1,\sigma_1}\|_{\mathrm{Lip}} = \frac{1}{\sigma_1\sqrt{2\pi}}.$$

Now, our aim is to replace in Theorem 4.1 the parameters (a_1, σ_1), which depend on ε, with $(0, v_1)$ like in Theorem 2.3. That is, we have the following:

Question. Assume that $\mathbf{E}X = \mathbf{E}Y = 0$, $\mathrm{Var}(X + Y) = 1$, and $L(F * G, \Phi) \le \varepsilon < 1$. Is it true that

$$L(F, \Phi_{v_1}) \le C\, \frac{\left(\log\log\frac{4}{\varepsilon}\right)^2}{\sqrt{\log\frac{1}{\varepsilon}}}$$

with some absolute constant C, where $v_1^2 = \mathrm{Var}(X)$?

In a sense, it is the question on the closeness of σ_1 to v_1 in the situation, where σ_1 is small. Indeed, using the triangle inequality, one can write

$$L(F, \Phi_{v_1}) \le L(F, \Phi_{a_1,\sigma_1}) + L(\Phi_{a_1,\sigma_1}, \Phi_{0,\sigma_1}) + L(\Phi_{\sigma_1}, \Phi_{v_1}).$$

Here, the first term may be estimated according to Theorem 4.1. For the second one, we have a trivial bound $L(\Phi_{a_1,\sigma_1}, \Phi_{0,\sigma_1}) \le |a_1|$, which follows from the definition of the Lévy metric. In turn, the parameter a_1 admits the bound, which was already used in the proof of Theorem 2.3, $|a_1| < \frac{1}{\sqrt{\log\frac{e}{\varepsilon}}}$. This bound behaves better than the one in Theorem 4.1, so we obtain:

Lemma 4.2. *If* $\mathbf{E}X = \mathbf{E}Y = 0$, $\mathrm{Var}(X+Y) = 1$, *and* $L(F*G, \Phi) \le \varepsilon < 1$, *then*

$$L(F, \Phi_{v_1}) \le C\, \frac{\left(\log\log\frac{4}{\varepsilon}\right)^2}{\sqrt{\log\frac{1}{\varepsilon}}} + L(\Phi_{\sigma_1}, \Phi_{v_1}).$$

Thus, we are reduced to estimating the distance $L(\Phi_{\sigma_1}, \Phi_{v_1})$, which in fact should be done in terms of $v_1^2 - \sigma_1^2$. The proof of the following elementary bound can be found in [3].

Lemma 4.3. *If* $v \ge \sigma \ge 0$, $v^2 - \sigma^2 \le 1$, *then* $L(\Phi_\sigma, \Phi_v)^2 \le (v^2 - \sigma^2)\,\log\frac{2}{v^2-\sigma^2}$.

Attempts to derive bounds on the distance $L(\Phi_\sigma, \Phi_v)$ by virtue of standard general relations, such as Zolotarev's Berry–Esseen-type estimate [25], lead to worse dependences of $\alpha^2 = v^2 - \sigma^2$.

In view of Lemmas 4.2 and 4.3, to proceed, one needs to bound $v_1^2 - \sigma_1^2$ in terms of ε. However, this does not seem to be possible in general without stronger hypotheses. Note that $v_1^2 - \sigma_1^2 = \int_{\{|x|>N\}} x^2\,dF(x) + a_1^2$. Hence, we need to deal with the quadratic tail function $\delta_X(T) = \int_{\{|x|>T\}} x^2\,dF(x) (T \ge 0)$, whose behavior at infinity will play an important role in the sequel.

Now, combining Lemmas 4.2 and 4.3, we obtain

$$L(F, \Phi_{v_1}) \le C\, \frac{\left(\log\log\frac{4}{\varepsilon}\right)^2}{\sqrt{\log\frac{1}{\varepsilon}}} + R\left(\delta_X(N) + a_1^2\right),$$

where $R(t) = \sqrt{t\log(2/t)}$. This function is non-negative and concave in the interval $0 \le t \le 2$, with $R(0) = 0$. Hence, it is subadditive in the sense that $R(\xi+\eta) \le R(\xi) + R(\eta)$, for all $\xi, \eta \ge 0$, $\xi + \eta \le 2$. Hence,

$$R\left(\delta_X(N) + a_1^2\right) \le R(\delta_X(N)) + R(a_1^2) = \left(\delta_X(N)\log\frac{2}{\delta_X(N)}\right)^{1/2} + \sqrt{a_1^2\log(2/a_1^2)}.$$

As we have noticed, $|a_1| \le A = \frac{1}{\sqrt{\log\frac{e}{\varepsilon}}}$, so $|a_1| \le 1$. Since $t \to t\log(e/t)$ is increasing on $[0, 1]$,

$$a_1^2\log\frac{2}{a_1^2} \le a_1^2\log\frac{e}{a_1^2} \le A^2\log\frac{e}{A^2} = \frac{1}{\log\frac{e}{\varepsilon}}\left(1 + \log\log\frac{e}{\varepsilon}\right).$$

Taking the square root of the right-hand side, we obtain a function which is majorized and absorbed by the bound of Theorem 4.1. As a result, we arrive at the following consequence of this theorem.

Theorem 4.4. *Assume independent random variables X and Y have distribution functions F and G with mean zero and with* $\mathrm{Var}(X+Y) = 1$. *If* $L(F*G, \Phi) \le \varepsilon < 1$, *then with some absolute constant C*

$$L(F, \Phi_{v_1}) \le C\, \frac{\left(\log\log\frac{4}{\varepsilon}\right)^2}{\sqrt{\log\frac{1}{\varepsilon}}} + \sqrt{\delta_X(N)\,\log(2/\delta_X(N))},$$

where $v_1 = \sqrt{\mathrm{Var}(X)}$, $N = 1 + \sqrt{2\log(1/\varepsilon)}$, *and* $\delta_X(N) = \int_{\{|x|>N\}} x^2\,dF(x)$.

It seems that in general it is not enough to know that $\mathrm{Var}(X) \le 1$ and $L(F * G, \Phi) \le \varepsilon < 1$, in order to judge the decay of the quadratic tail function $\delta_X(T)$ as $T \to \infty$. So, some additional properties should be involved. As we will see, the entropic distance perfectly suits this idea, so that one can start with the entropic assumption $D(X + Y) \le \varepsilon$.

5. Application of Sapogov-type results to Gaussian regularization

In this section we consider the stability problem in Cramer's theorem for the regularized distributions with respect to the total variation norm. As a basic tool, we use Theorem 2.3.

Thus, let X and Y be independent random variables with distribution functions F and G, and with variances $\mathrm{Var}(X) = v_1^2$, $\mathrm{Var}(Y) = v_2^2 (v_1, v_2 > 0,\ v_1^2 + v_2^2 = 1)$, so that $X + Y$ has variance 1. What is not important (and is assumed for simplicity of notations, only), let both X and Y have mean zero. As we know from Theorem 2.3, the main stability result asserts that if $\|F * G - \Phi\| \le \varepsilon < 1$, then

$$\|F - \Phi_{v_1}\| \le \frac{Cm(v_1, \varepsilon)}{v_1 \sqrt{\log \frac{1}{\varepsilon}}}, \qquad \|G - \Phi_{v_2}\| \le \frac{Cm(v_2, \varepsilon)}{v_2 \sqrt{\log \frac{1}{\varepsilon}}}$$

for some absolute constant C. Here, as before $m(v, \varepsilon) = \min\{\frac{1}{\sqrt{v}}, \log\log \frac{e^e}{\varepsilon}\}$, $v > 0, 0 < \varepsilon \le 1$.

On the other hand, such a statement – even in the case of equal variances – is no longer true for the total variation norm. So, it is natural to use the Gaussian regularizations $X_\sigma = X + \sigma Z$, $Y_\sigma = Y + \sigma Z$, where $Z \sim N(0, 1)$ is independent of X and Y, and where σ is a (small) positive parameter. For definiteness, we assume that $0 < \sigma \le 1$. Note that

$$\mathrm{Var}(X_\sigma) = v_1^2 + \sigma^2, \qquad \mathrm{Var}(Y_\sigma) = v_2^2 + \sigma^2 \quad \text{and} \quad \mathrm{Var}(X_\sigma + Y_\sigma) = 1 + 2\sigma^2.$$

Denote by F_σ and G_σ the distributions of X_σ and Y_σ, respectively. Assume $X_\sigma + Y_\sigma$ is almost normal in the sense of the total variation norm and hence in the Kolmogorov distance, namely,

$$\|F_\sigma * G_\sigma - N(0, 1 + 2\sigma^2)\| \le \frac{1}{2} \qquad \|F_\sigma * G_\sigma - N(0, 1 + 2\sigma^2)\|_{\mathrm{TV}} \le \varepsilon \le 1.$$

Note that $X_\sigma + Y_\sigma = (X + Y) + \sigma\sqrt{2}\, Z$ represents the Gaussian regularization of the sum $X + Y$ with parameter $\sigma\sqrt{2}$. One may also write $X_\sigma + Y_\sigma = X + (Y + \sigma\sqrt{2}\, Z)$, or equivalently,

$$\frac{X_\sigma + Y_\sigma}{\sqrt{1 + 2\sigma^2}} = X' + Y', \quad \text{where } X' = \frac{X}{\sqrt{1 + 2\sigma^2}}, \qquad Y' = \frac{Y + \sigma\sqrt{2}\, Z}{\sqrt{1 + 2\sigma^2}}.$$

Thus, we are in position to apply Theorem 2.3 to the distributions of the random variables X' and Y' with variances $v_1'^2 = \frac{v_1^2}{1+2\sigma^2}$ and $v_2'^2 = \frac{v_2^2 + 2\sigma^2}{1+2\sigma^2}$. Using $1 + 2\sigma^2 \le 3$, it gives

$$\|F - \Phi_{v_1}\| \le \frac{Cm(v_1', \varepsilon)}{v_1' \sqrt{\log \frac{1}{\varepsilon}}} \le \frac{3Cm(v_1, \varepsilon)}{v_1 \sqrt{\log \frac{1}{\varepsilon}}}.$$

Now, we apply Proposition B.2(b) to the distributions F and $G = \Phi_{v_1}$ with $B = v_1$ and get

$$\left\| F_\sigma - N(0, v_1^2 + \sigma^2) \right\|_{\mathrm{TV}} \le \frac{4v_1}{\sigma} \|F - \Phi_{v_1}\|^{1/2} \le \frac{4v_1}{\sigma} \frac{\sqrt{3Cm(v_1, \varepsilon)}}{v_1^{1/2} \left(\log \frac{1}{\varepsilon}\right)^{1/4}}.$$

One may simplify this bound by using $v_1 \sqrt{m(v_1, \varepsilon)} \le \sqrt{v_1}$, and then we may conclude:

Theorem 5.1. *Let F and G be distribution functions with mean zero and variances v_1^2, v_2^2, respectively, such that $v_1^2 + v_2^2 = 1$. Let $0 < \sigma \leq 1$. If the regularized distributions satisfy*

$$\frac{1}{2}\, \big\| F_\sigma * G_\sigma - N(0, 1 + 2\sigma^2) \big\|_{\mathrm{TV}} \leq \varepsilon \leq 1,$$

then with some absolute constant C

$$\big\| F_\sigma - N(0, v_1^2 + \sigma^2) \big\|_{\mathrm{TV}} \leq \frac{C}{\sigma} \left(\frac{1}{\log \frac{1}{\varepsilon}} \right)^{1/4},$$

$$\big\| G_\sigma - N(0, v_2^2 + \sigma^2) \big\|_{\mathrm{TV}} \leq \frac{C}{\sigma} \left(\frac{1}{\log \frac{1}{\varepsilon}} \right)^{1/4}.$$

6. Control of tails and entropic Chebyshev-type inequality

One of our further aims is to find an entropic version of the Sapogov stability theorem for regularized distributions. As part of the problem, we need to bound the quadratic tail function $\delta_X(T) = \mathbf{E} X^2\, 1_{\{|X| \geq T\}}$ quantitatively in terms of the entropic distance $D(X)$. Thus, assume a random variable X has mean zero and variance $\mathrm{Var}(X) = 1$, with a finite distance to the standard normal law

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

where p is density of X and φ is the density of $N(0, 1)$. One can also write another representation, $D(X) = \mathrm{Ent}_\gamma(f)$, where $f = \frac{p}{\varphi}$, with respect to the standard Gaussian measure γ on the real line. Let us recall that the entropy functional

$$\mathrm{Ent}_\mu(f) = \mathbf{E}_\mu f \log f - \mathbf{E}_\mu f \log \mathbf{E}_\mu f$$

is well-defined for any measurable function $f \geq 0$ on an abstract probability space (Ω, μ), where $\mathbf{E}_\mu$ stands for the expectation (integral) with respect to μ.

We are going to involve a variational formula for this functional (cf. e.g. [12]): For all measurable functions $f \geq 0$ and g on Ω, such that $\mathrm{Ent}_\mu(f)$ and $\mathbf{E}_\mu\, e^g$ are finite,

$$\mathbf{E}_\mu f g \leq \mathrm{Ent}_\mu(f) + \mathbf{E}_\mu f \log \mathbf{E}_\mu\, e^g.$$

Applying it on $\Omega = \mathbf{R}$ with $\mu = \gamma$ and $f = \frac{p}{\varphi}$, we notice that $\mathbf{E}_\mu f = 1$ and get that

$$\int_{-\infty}^{\infty} p(x)\, g(x)\, dx \ \leq \ D(X) + \log \int_{-\infty}^{\infty} e^{g(x)}\, \varphi(x)\, dx.$$

Take $g(x) = \frac{\alpha}{2}\, x^2\, 1_{\{|x| \geq T\}}$ with a parameter $\alpha \in (0, 1)$. Then,

$$\int_{-\infty}^{\infty} e^{g(x)}\, \varphi(x)\, dx = \gamma[-T, T] + \frac{2}{\sqrt{1-\alpha}} \left(1 - \Phi\left(T\sqrt{1-\alpha}\right)\right).$$

Using $\gamma[-T, T] < 1$ and the inequality $\log(1 + t) \leq t$, we obtain that

$$\frac{1}{2}\, \delta_X(T) \leq \frac{1}{\alpha}\, D(X) + \frac{2}{\alpha\sqrt{1-\alpha}} \left(1 - \Phi\left(T\sqrt{1-\alpha}\right)\right).$$

To further estimate the right-hand side, we apply the bound $1 - \Phi(t) \leq \varphi(t)/t$, which leads to

$$\frac{1}{2}\,\delta_X(T) \leq \frac{1}{\alpha}\,D(X) + \frac{2}{\sqrt{2\pi}}\,\frac{1}{T\alpha(1-\alpha)}\,e^{-(1-\alpha)\,T^2/2}. \tag{6.1}$$

Choosing just $\alpha = 1/2$, we get

$$\frac{1}{2}\,\delta_X(T) \leq 2D(X) + \frac{8}{T\sqrt{2\pi}}\,e^{-T^2/4} \leq 2D(X) + 2\,e^{-T^2/4},$$

where the last bounds are fulfilled for $T \geq 4/\sqrt{2\pi}$. For the remaining T the obtained inequality is fulfilled automatically, since then $2e^{-T^2/4} \geq 2e^{-4/2\pi} > 1$, while $\frac{1}{2}\,\delta_X(T) \leq \frac{1}{2}\,\mathbf{E}X^2 = \frac{1}{2}$. Thus, we have proved the following:

Proposition 6.1. *If X is a random variable with $\mathbf{E}X = 0$ and $\mathrm{Var}(X) = 1$, having density $p(x)$, then for all $T > 0$,*

$$\int_{\{|x| \geq T\}} x^2\,p(x)\,dx \leq 4D(X) + 4\,e^{-T^2/4}.$$

In particular, the above integral does not exceed $8D(X)$ for $T = 2\sqrt{\log^+(1/D(X))}$.

The choice $\alpha = 2/T^2$ in (6.1) leads to a better asymptotic in T, and then we also have:

Proposition 6.2. *If X is a random variable with $\mathbf{E}X = 0$ and $\mathrm{Var}(X) = 1$, having density $p(x)$, then for all $T \geq 2$,*

$$\int_{\{|x| \geq T\}} x^2\,p(x)\,dx \leq T^2 D(X) + 6T\,e^{-T^2/2}.$$

In the Gaussian case $X = Z$ this gives an asymptotically correct bound for $T \to \infty$ (up to a factor). Note as well that in the non-Gaussian case, from Proposition 6.1 we obtain an entropic Chebyshev-type inequality

$$\mathbf{P}\left\{|X| \geq 2\sqrt{\log(1/D(X))}\right\} \leq \frac{2D(X)}{\log(1/D(X))} \quad (D(X) < 1).$$

Finally, let us give a more flexible variant of Proposition 6.1 with an arbitrary variance $B^2 = \mathrm{Var}(X)$ $(B > 0)$, but still with mean zero. Applying the obtained statements to the random variable X/B and replacing the variable T with T/B, we then get that

$$\frac{1}{B^2}\int_{\{|x| \geq T\}} x^2\,p(x)\,dx \leq 4D(X) + 4\,e^{-T^2/4B^2}.$$

7. Entropic control of tails for sums of independent summands

We apply Proposition 6.1 in the following situation. Assume we have two independent random variables X and Y with mean zero, but perhaps with different variances $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$. Assume they have densities. The question is: Can we bound the tail functions δ_X and δ_Y in terms of $D(X + Y)$, rather than in terms of $D(X)$ and $D(Y)$? In case $\mathrm{Var}(X + Y) = 1$, by

Proposition 6.1, applied to the sum $X + Y$,

$$\delta_{X+Y}(T) = \mathbf{E}\,(X+Y)^2\, 1_{\{|X+Y| \geq T\}} \leq 4\,D(X+Y) + 4\,e^{-T^2/4}. \tag{7.1}$$

Hence, to answer the question, it would be sufficient to bound from below the tail functions δ_{X+Y} in terms of δ_X and δ_Y.

Assume for a while that $\mathrm{Var}(X+Y) = 1/2$. In particular, $\mathrm{Var}(Y) \leq 1/2$, and according to the usual Chebyshev's inequality, $\mathbf{P}\{Y \geq -1\} \geq \frac{1}{2}$. Hence, for all $T \geq 0$,

$$\mathbf{E}\,(X+Y)^2\, 1_{\{X+Y \geq T\}} \geq \mathbf{E}\,(X+Y)^2\, 1_{\{X \geq T+1,\, Y \geq -1\}}$$
$$\geq \mathbf{E}\,(X-1)^2\, 1_{\{X \geq T+1,\, Y \geq -1\}} \geq \frac{1}{2}\,\mathbf{E}\,(X-1)^2\, 1_{\{X \geq T+1\}}.$$

If $X \geq T+1 \geq 4$, then clearly $(X-1)^2 \geq \frac{1}{2}\,X^2$, hence, $\mathbf{E}\,(X-1)^2\, 1_{\{X \geq T+1\}} \geq \frac{1}{2}\,\mathbf{E}\,X^2\, 1_{\{X \geq T+1\}}$. With a similar bound for the range $X \leq -(T+1)$, we get

$$\delta_{X+Y}(T) \geq \frac{1}{4}\,\delta_X(T+1), \quad T \geq 3. \tag{7.2}$$

Now, change $T+1$ with T (assuming that $T \geq 4$) and apply (7.1) to $\sqrt{2}\,(X+Y)$. Together with (7.2) it gives $\frac{1}{4}\,\delta_{\sqrt{2}X}(T) \leq 4\,D\big(\sqrt{2}\,(X+Y)\big) + 4\,e^{-(T-1)^2/4}$. But the entropic distance to the normal is invariant under rescaling of coordinates, i.e., $D(\sqrt{2}\,(X+Y)) = D(X+Y)$. Since also $\delta_{\sqrt{2}X}(T) = 2\,\delta_X(T/\sqrt{2})$, we obtain that

$$\delta_X(T/\sqrt{2}) \leq 8\,D(X+Y) + 8\,e^{-(T-1)^2/4},$$

provided that $T \geq 4$. Simplifying by $e^{-(T-1)^2/4} \leq e^{-T^2/8}$ (valid for $T \geq 4$), and then replacing T with $T\sqrt{2}$, we arrive at

$$\delta_X(T) \leq 8\,D(X+Y) + 8\,e^{-T^2/4}, \quad T \geq 4/\sqrt{2}.$$

Finally, to involve the values $0 \leq T \leq 4/\sqrt{2}$, just use $e^2 < 8$, so that the above inequality holds automatically for this range: $\delta_X(T) \leq \mathrm{Var}(X) \leq 1 < 8\,e^{-T^2/4}$. Moreover, in order to allow an arbitrary variance $\mathrm{Var}(X+Y) = B^2$ $(B > 0)$, the above estimate should be applied to $X/B\sqrt{2}$ and $Y/B\sqrt{2}$ with T replaced by $T/B\sqrt{2}$. Then it takes the form

$$\frac{1}{2B^2}\,\delta_X(T) \leq 8\,D(X+Y) + 8\,e^{-T^2/8B^2}.$$

We can summarize.

Proposition 7.1. *Let X and Y be independent random variables with mean zero and with* $\mathrm{Var}(X+Y) = B^2$ $(B > 0)$. *Assume X has a density p. Then, for all $T \geq 0$,*

$$\frac{1}{B^2}\int_{\{|x| \geq T\}} x^2\, p(x)\,dx \leq 16\,D(X+Y) + 16\,e^{-T^2/8B^2}.$$

8. Stability for Lévy distance under entropic hypothesis

Now we can return to the variant of the Chistyakov–Golinskii result, as in Theorem 4.4. Let the independent random variables X and Y have mean zero, with $\mathrm{Var}(X+Y) = 1$, and denote by F and G their distribution functions. Also assume X has a density p. In order to control the term $\delta_X(N)$ in Theorem 4.4, we are going to impose the stronger condition $D(X+Y) \le 2\varepsilon$. Using Pinsker's inequality, this yields bounds for the total variation and Kolmogorov distances

$$\|F * G - \Phi\| \le \frac{1}{2} \qquad \|F * G - \Phi\|_{\rm TV} \le \frac{1}{2}\sqrt{2D(X+Y)} \le \sqrt{\varepsilon} = \varepsilon'.$$

Hence, the assumption of Theorem 4.4 is fulfilled, whenever $\varepsilon < 1$. As for the conclusion, first apply Proposition 7.1 with $B = 1$, which gives

$$\delta_X(T) = \int_{\{|x|\ge T\}} x^2\, p(x)\, dx \le 16\, D(X+Y) + 16\, e^{-T^2/8} \le 16\, \varepsilon + 16\, e^{-T^2/8}.$$

In our situation, $N = 1+\sqrt{2\log(1/\varepsilon')} = 1+\sqrt{\log(1/\varepsilon)}$, so, $\delta_X(N) \le 16\,\varepsilon + 16\, e^{-N^2/8} \le C\varepsilon^{1/8}$. Thus, we arrive at:

Proposition 8.1. *Let the independent random variables X and Y have mean zero, with* $\mathrm{Var}(X+Y) = 1$*, and assume that X has a density with distribution function F. If $D(X+Y) \le 2\varepsilon < 2$, then*

$$L(F, \Phi_{v_1}) \le C\, \frac{\left(\log\log \frac{4}{\varepsilon}\right)^2}{\sqrt{\log \frac{1}{\varepsilon}}},$$

where $v_1 = \sqrt{\mathrm{Var}(X)}$ and C is an absolute constant.

In general, in the conclusion one cannot replace the Lévy distance $L(F, \Phi_{v_1})$ with $D(X)$. However, this is indeed possible for regularized distributions, as we will see in the next sections.

9. Entropic distance and uniform deviation of densities

Let X and Y be independent random variables with mean zero, finite variances, and assume X has a bounded density p. Our next aim is to estimate the entropic distance to the normal, $D(X)$, in terms of $D(X+Y)$ and the uniform deviation of p above the normal density

$$\Delta(X) = \operatorname{ess\,sup}_x\, (p(x) - \varphi_v(x)),$$

where $v^2 = \mathrm{Var}(X)$ and φ_v stands for the density of the normal law $N(0, v^2)$.

For a while, assume that $\mathrm{Var}(X) = 1$. Proposition C.2 gives the preliminary estimate

$$D(X) \le \Delta(X)\left[\sqrt{2\pi} + 2T + 2T\, \log\left(1 + \Delta(X)\sqrt{2\pi}\, e^{T^2/2}\right)\right] + \frac{1}{2}\,\delta_X(T),$$

involving the quadratic tail function $\delta_X(T)$. In the general situation one cannot say anything definite about the decay of this function. However, it can be bounded in terms of $D(X+Y)$ by virtue of Proposition 7.1: we know that, for all $T \ge 0$,

$$\frac{1}{2B^2}\,\delta_X(T) \le 8\, D(X+Y) + 8\, e^{-T^2/8B^2},$$

where $B^2 = \mathrm{Var}(X+Y) = 1 + \mathrm{Var}(Y)$. So, combining the two estimates yields

$$D(X) \le 8B^2\, D(X+Y) + 8B^2\, e^{-T^2/8B^2} + \Delta\left[\sqrt{2\pi} + 2T + 2T\, \log\big(1 + \Delta\sqrt{2\pi}\, e^{T^2/2}\big)\right], \quad \text{where } \Delta = \Delta(X).$$

First assume $\Delta \le 1$ and apply the above with $T^2 = 8B^2 \log \frac{1}{\Delta}$. Then $8B^2\, e^{-T^2/8B^2} = 8B^2\, \Delta$, and putting $\beta = 4B^2 - 1 \ge 3$, we also have

$$\log\big(1 + \Delta\sqrt{2\pi}\, e^{T^2/2}\big) = \log\Big(1 + \Delta^{-\beta}\sqrt{2\pi}\Big) = \beta \log\Big(1 + \Delta^{-\beta}\sqrt{2\pi}\Big)^{1/\beta}$$
$$< \beta \log\bigg(1 + \frac{(2\pi)^{1/2\beta}}{\Delta}\bigg) < \beta \log\Big(1 + \frac{2}{\Delta}\Big).$$

Collecting all the terms and using $B \ge 1$, we are led to the estimate of the form

$$D(X) \le 8B^2\, D(X+Y) + C B^3\, \Delta \log^{3/2}\Big(2 + \frac{1}{\Delta}\Big),$$

where $C > 0$ is an absolute constant. It holds also in case $\Delta > 1$ in view of the logarithmic bound of Proposition C.1,

$$D(X) \le \log\Big(1 + \Delta\sqrt{2\pi}\Big) + \frac{1}{2}.$$

Therefore, the obtained bound holds true without any restriction on Δ.

Now, to relax the variance assumption, assume $\mathrm{Var}(X) = v_1^2$, $\mathrm{Var}(Y) = v_2^2$ $(v_1, v_2 > 0)$, and without loss of generality, let $\mathrm{Var}(X+Y) = v_1^2 + v_2^2 = 1$. Apply the above to $X' = \frac{X}{v_1}$, $Y' = \frac{Y}{v_1}$. Then, $B^2 = 1/v_1^2$ and $\Delta(X') = v_1\, \Delta(X)$, so with some absolute constant $c > 0$,

$$c\, v_1^2\, D(X) \le D(X+Y) + \Delta(X) \log^{3/2}\Big(2 + \frac{1}{v_1 \Delta(X)}\Big).$$

As a result, we arrive at:

Proposition 9.1. *Let X, Y be independent random variables with mean zero,* $\mathrm{Var}(X+Y) = 1$, *and such that X has a bounded density. Then, with some absolute constant $c > 0$,*

$$c\, \mathrm{Var}(X)\, D(X) \le D(X+Y) + \Delta(X) \log^{3/2}\bigg(2 + \frac{1}{\sqrt{\mathrm{Var}(X)}\, \Delta(X)}\bigg).$$

Replacing the role of X and Y, and adding the two inequalities, we also have as corollary:

Proposition 9.2. *Let X, Y be independent random variables with mean zero and positive variances $v_1^2 = \mathrm{Var}(X)$, $v_2^2 = \mathrm{Var}(Y)$, such that $v_1^2 + v_2^2 = 1$, and both with densities. Then, with some absolute constant $c > 0$,*

$$c\, (v_1^2\, D(X) + v_2^2\, D(Y)) \le D(X+Y) + \Delta(X) \log^{3/2}\Big(2 + \frac{1}{v_1 \Delta(X)}\Big)$$
$$+ \Delta(Y) \log^{3/2}\Big(2 + \frac{1}{v_2 \Delta(Y)}\Big).$$

This inequality may be viewed as the inverse to the general property of the entropic distance, which we mentioned before, namely, $v_1^2\, D(X) + v_2^2\, D(Y) \ge D(X+Y)$, under the normalization assumption $v_1^2 + v_2^2 = 1$. Let us also state separately Proposition 9.1 in the particular case of equal unit variances, keeping the explicit constant $8B^2 = 16$ in front of $D(X+Y)$.

Proposition 9.3. *Let X, Y be independent random variables with mean zero and variances* $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$*, and such that X has a density. Then, with some absolute constant C*

$$D(X) \le 16\, D(X+Y) + C\, \Delta(X) \log^{3/2}\Big(2 + \frac{1}{\Delta(X)}\Big).$$

One may simplify the right-hand side for small values of $\Delta(X)$ and get a slightly weaker inequality $D(X) \le 16\, D(X+Y) + C_\alpha\, \Delta(X)^\alpha$, $0 < \alpha < 1$, where the constants C_α depend on α, only. For large values of $\Delta(X)$, the above inequality holds, as well, in view of the logarithmic bound of Proposition C.1.

10. The case of equal variances

We are prepared to derive an entropic variant of Sapogov-type stability theorem for regularized distributions. That is, we are going to estimate $D(X_\sigma)$ and $D(Y_\sigma)$ in terms of $D(X_\sigma + Y_\sigma)$ for two independent random variables X and Y with distribution functions F and G, by involving a small "smoothing" parameter $\sigma > 0$. It will not be important whether or not they have densities. Since it will not be important for the final statements, let X and Y have mean zero. Recall that, given $\sigma > 0$, the regularized random variables are defined by $X_\sigma = X + \sigma Z, Y_\sigma = Y + \sigma Z$, where Z is independent of X and Y, and has a standard normal density φ. The distributions of X_σ, Y_σ are denoted F_σ, G_σ, with densities p_σ, q_σ.

In this section, we consider the case of equal variances, say, $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$. Put $\sigma_1 = \sqrt{1+\sigma^2}$, $\sigma_2 = \sqrt{1+2\sigma^2}$. Since $\mathrm{Var}(X_\sigma) = \mathrm{Var}(Y_\sigma) = \sigma_1^2$, the corresponding entropic distances are given by

$$D(X_\sigma) = h(\sigma_1 Z) - h(X_\sigma) = \int_{-\infty}^{\infty} p_\sigma(x)\, \log \frac{p_\sigma(x)}{\varphi_{\sigma_1}(x)}\, dx,$$

and similarly for Y_σ, where, as before, φ_v represents the density of $N(0, v^2)$. Assume that $D(X_\sigma + Y_\sigma)$ is small in the sense that $D(X_\sigma + Y_\sigma) \le 2\varepsilon < 2$. According to Pinsker's inequality, this yields bounds for the total variation and Kolmogorov distances

$$\|F_\sigma * G_\sigma - \Phi_{\sigma_2}\| \le \frac{1}{2}\, \|F_\sigma * G_\sigma - \Phi_{\sigma_2}\|_{\mathrm{TV}} \le \sqrt{\varepsilon} < 1.$$

In the sequel, let $0 < \sigma \le 1$. This guarantees that the ratio of variances of the components in the convolution $F_\sigma * G_\sigma = F * (G * \Phi_{\sigma\sqrt{2}})$ is bounded away from zero by an absolute constant, so that we can apply Theorem 2.3. Namely, it gives that $\|F - \Phi\| \le C \log^{-1/2}(\frac{1}{\varepsilon})$, and similarly for G. (Note that raising ε to any positive power does not change the above estimate.) Applying Proposition B.1(a), when one of the distributions is normal, we get

$$\Delta(X_\sigma) = \sup_x\, (p_\sigma(x) - \varphi_{\sigma_1}(x)) \le \frac{1}{\sigma}\, \|F - \Phi\| \le \frac{C}{\sigma\sqrt{\log\frac{1}{\varepsilon}}}.$$

We are in position to apply Proposition 9.3 to the random variables X_σ/σ_1, Y_σ/σ_1. It gives

$$D(X_\sigma) \leq 16\, D(X_\sigma + Y_\sigma) + C\, \varDelta(X_\sigma) \log^{3/2}\Big(2 + \frac{1}{\varDelta(X_\sigma)}\Big)$$
$$\leq 32\,\varepsilon + C'\, \frac{\log^{3/2}\big(2 + \sigma\sqrt{\log\frac{1}{\varepsilon}}\big)}{\sigma\sqrt{\log\frac{1}{\varepsilon}}},$$

where C' is an absolute constant. In the last expression the second term dominates the first one, and at this point, the assumption on the means may be removed. We arrive at:

Proposition 10.1. *Let X and Y be independent random variables with variance one. Given $0 < \varepsilon < 1$ and $0 < \sigma \leq 1$, the regularized random variables X_σ and Y_σ satisfy*

$$D(X_\sigma + Y_\sigma) \leq 2\varepsilon \Rightarrow D(X_\sigma) + D(Y_\sigma) \leq C\, \frac{\log^{3/2}\big(2 + \sigma\sqrt{\log\frac{1}{\varepsilon}}\big)}{\sigma\sqrt{\log\frac{1}{\varepsilon}}}, \tag{10.1}$$

where C is an absolute constant.

Note that all entropic distances in (10.1) do not change when adding constants to X and Y, which allows us to remove the mean zero assumption. This statement may be formulated equivalently by solving the above inequality with respect to ε. The function $u(x) = \frac{x}{\log^{3/2}(2+x)}$ is increasing in $x \geq 0$, and, for any $a \geq 0$, $u(x) \leq a \Rightarrow x \leq 8\,a \log^{3/2}(2 + a)$. Hence, assuming $D(X_\sigma + Y_\sigma) \leq 1$, we obtain from (10.1) that

$$\sigma\sqrt{\log\frac{1}{\varepsilon}} \leq \frac{8C}{D}\, \log^{3/2}(2 + C/D) \leq \frac{C'}{D}\, \log^{3/2}(2 + 1/D)$$

with some absolute constant C', where $D = D(X_\sigma) + D(Y_\sigma)$. As a result,

$$D(X_\sigma + Y_\sigma) \geq \exp\Big\{-\frac{C'^2 \log^3(2 + 1/D)}{\sigma^2 D^2}\Big\}.$$

Note also that this inequality is fulfilled automatically, if $D(X_\sigma + Y_\sigma) \geq 1$. Thus, we get:

Proposition 10.2. *Let X, Y be independent random variables with* $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$. *Given $0 < \sigma \leq 1$, the regularized random variables X_σ and Y_σ satisfy*

$$D(X_\sigma + Y_\sigma) \geq \exp\Big\{-\frac{C\, \log^3(2 + 1/D)}{\sigma^2 D^2}\Big\},$$

where $D = D(X_\sigma) + D(Y_\sigma)$ and $C > 0$ is an absolute constant.

11. Proof of Theorem 1.1

Now let us consider the case of arbitrary variances $\mathrm{Var}(X) = v_1^2$, $\mathrm{Var}(Y) = v_2^2 (v_1, v_2 \geq 0)$. For normalization reasons, let $v_1^2 + v_2^2 = 1$. Then

$$\mathrm{Var}(X_\sigma) = v_1^2 + \sigma^2, \qquad \mathrm{Var}(Y_\sigma) = v_2^2 + \sigma^2, \qquad \mathrm{Var}(X_\sigma + Y_\sigma) = \sigma_2^2,$$

where $\sigma_2 = \sqrt{1 + 2\sigma^2}$. As before, we assume that both X and Y have mean zero, although this will not be important for the final conclusion.

Again, we start with the hypothesis $D(X_\sigma + Y_\sigma) \leq 2\varepsilon < 2$ and apply Pinsker's inequality:

$$\|F_\sigma * G_\sigma - \Phi_{\sigma_2}\| \leq \frac{1}{2}\,\|F_\sigma * G_\sigma - \Phi_{\sigma_2}\|_{\rm TV} \leq \sqrt{\varepsilon} < 1.$$

For $0 < \sigma \leq 1$, write $F_\sigma * G_\sigma = F * (G * \Phi_{\sigma\sqrt{2}})$. Now, the ratio of variances of the components in the convolution, $\frac{v_1^2}{1+2\sigma^2}$, may not be bounded away from zero, since v_1 is allowed to be small. Hence, the application of Theorem 2.3 will only give $\|F - \Phi_{v_1}\| \leq \frac{Cm(v_1,\varepsilon)}{v_1\sqrt{\log\frac{1}{\varepsilon}}}$ and similarly for G. The appearance of v_1 on the right is however not desirable. So, it is better to involve the Lévy distance, which is more appropriate in such a situation. Consider the random variables

$$X' = \frac{X}{\sqrt{1+2\sigma^2}}, \qquad Y' = \frac{Y + \sigma\sqrt{2}Z}{\sqrt{1+2\sigma^2}},$$

so that $\mathrm{Var}(X'+Y') = 1$, and denote by F', G' their distribution functions. Since the Kolmogorov distance does not change after rescaling of the coordinates, we still have

$$L(F' * G', \Phi) \leq \|F' * G' - \Phi\| = \|F_\sigma * G_\sigma - \Phi_{\sigma_2}\| \leq \sqrt{\varepsilon} < 1.$$

In this situation, we may apply Proposition 8.1 to the couple (F', G'). It gives that

$$L(F', \Phi_{v_1'}) \leq C\left(\log\log\frac{4}{\varepsilon}\right)^2\left(\log\frac{1}{\varepsilon}\right)^{-1/2}$$

with some absolute constant C, where $v_1' = \sqrt{\mathrm{Var}(X')} = \frac{v_1}{\sqrt{1+2\sigma^2}}$. Since $v_1' \leq v_1 \leq \sqrt{3}v_1'$, we have a similar conclusion about the original distribution functions, i.e. $L(F, \Phi_{v_1}) \leq C\,(\log\log\frac{4}{\varepsilon})^2\,(\log\frac{1}{\varepsilon})^{-1/2}$. Now we use Proposition B.3 (applied when one of the distributions is normal), which for $\sigma \leq 1$ gives $\Delta(X_\sigma) \leq \frac{3}{2\sigma^2}\,L(F, \Phi_{v_1})$, and similarly for Y. Hence,

$$\Delta(X_\sigma) \leq C\,\frac{\left(\log\log\frac{4}{\varepsilon}\right)^2}{\sigma^2\sqrt{\log\frac{1}{\varepsilon}}}, \qquad \Delta(Y_\sigma) \leq C\,\frac{\left(\log\log\frac{4}{\varepsilon}\right)^2}{\sigma^2\sqrt{\log\frac{1}{\varepsilon}}}. \tag{11.1}$$

We are now in a position to apply Proposition 9.2 to the random variables $X_\sigma' = X_\sigma/\sqrt{1+\sigma^2}$, $Y_\sigma' = Y_\sigma/\sqrt{1+\sigma^2}$, which ensures that with some absolute constant $c > 0$

$$c\,(v_1(\sigma)^2\,D(X_\sigma) + v_2(\sigma)^2\,D(Y_\sigma)) \leq D(X_\sigma + Y_\sigma) + \Delta(X_\sigma)\log^{3/2}\Big(2 + \frac{1}{v_1(\sigma)\Delta(X_\sigma)}\Big)$$
$$+ \Delta(Y_\sigma)\log^{3/2}\Big(2 + \frac{1}{v_2(\sigma)\Delta(Y_\sigma)}\Big),$$

where $v_1(\sigma)^2 = \mathrm{Var}(X'_\sigma) = \frac{v_1^2+\sigma^2}{1+\sigma^2}$ and $v_2(\sigma)^2 = \mathrm{Var}(Y'_\sigma) = \frac{v_2^2+\sigma^2}{1+\sigma^2}$ ($v_1(\sigma), v_2(\sigma) \geq 0$). Note that $v_1(\sigma) \geq \sigma/\sqrt{2}$. Applying the bounds in (11.1), we obtain that

$$c\,(v_1(\sigma)^2\, D(X_\sigma) + v_2(\sigma)^2\, D(Y_\sigma)) \leq D(X_\sigma + Y_\sigma) + \frac{\left(\log\log\frac{4}{\varepsilon}\right)^2}{\sigma^2\sqrt{\log\frac{1}{\varepsilon}}}\,\log^{3/2}\Bigg(2 + \frac{\sigma\sqrt{\log\frac{1}{\varepsilon}}}{\left(\log\log\frac{4}{\varepsilon}\right)^2}\Bigg)$$

with some other absolute constant $c > 0$. Here, $D(X_\sigma + Y_\sigma) \leq 2\varepsilon$, which is dominated by the last expression, and we arrive at:

Proposition 11.1. *Let X, Y be independent random variables with $\mathrm{Var}(X + Y) = 1$. Given $0 < \sigma \leq 1$, if the regularized random variables X_σ, Y_σ satisfy $D(X_\sigma + Y_\sigma) \leq 2\varepsilon < 2$, then with some absolute constant C*

$$\mathrm{Var}(X_\sigma)\, D(X_\sigma) + \mathrm{Var}(Y_\sigma)\, D(Y_\sigma) \leq C\,\frac{\left(\log\log\frac{4}{\varepsilon}\right)^2}{\sigma^2\sqrt{\log\frac{1}{\varepsilon}}}\,\log^{3/2}\Bigg(2 + \frac{\sigma\sqrt{\log\frac{1}{\varepsilon}}}{\left(\log\log\frac{4}{\varepsilon}\right)^2}\Bigg). \tag{11.2}$$

It remains to solve this inequality with respect to ε. Denote by D' the left-hand side of (11.2) and let $D = \sigma^2 D'$. Assuming that $D(X_\sigma + Y_\sigma) < 2$ and arguing as in the proof of Proposition 10.2, we get $\frac{\sigma\sqrt{\log\frac{1}{\varepsilon}}}{(\log\log\frac{4}{\varepsilon})^2} \leq \frac{8C}{\sigma D'}\log^{3/2}(2 + C/D')$, hence $\frac{\log\frac{1}{\varepsilon}}{(\log\log\frac{4}{\varepsilon})^4} \leq A \equiv \frac{C'}{D^2}\log^3(2 + 1/D)$ with some absolute constant C'. The latter inequality implies with some absolute constants

$$\log\frac{1}{\varepsilon} \leq C'' A\log^4(2 + A) \leq \frac{C'''}{D^2}\log^7(2 + 1/D),$$

and we arrive at the inequality of Theorem 1.1 (which holds automatically, if $D(X_\sigma + Y_\sigma) \geq 1$).

Acknowledgment

This research was partially supported by NSF grant DMS-1612961, the Humboldt Foundation and SFB 701 at Bielefeld University.

Appendix A. General bounds for distances between distribution functions

Here we collect a few elementary and basically known relations for classical metrics, introduced at the beginning of Section 2. Let F and G be arbitrary distribution functions of some random variables X and Y. First of all, the Lévy, Kolmogorov, and the total variation distances are connected by the chain of the inequalities

$$0 \leq L(F, G) \leq \|F - G\| \leq \frac{1}{2}\,\|F - G\|_{\mathrm{TV}} \leq 1.$$

As for the Kantorovich–Rubinshtein distance, there is the following well-known bound.

Proposition A.1. *We have $L(F, G) \leq W_1(F, G)^{1/2}$.*

We also have:

Proposition A.2. *If* $\int_{-\infty}^{\infty} x^2\,dF(x) \le B^2$ *and* $\int_{-\infty}^{\infty} x^2\,dG(x) \le B^2$ $(B \ge 0)$, *then*

(a) $W_1(F,G) \le 2L(F,G) + 4B\,L(F,G)^{1/2}$;
(b) $W_1(F,G) \le 4B\,\|F-G\|^{1/2}$.

Appendix B. Relations for distances between regularized distributions

Now, let us turn to the regularized random variables $X_\sigma = X + \sigma Z$, $Y_\sigma = Y + \sigma Z$, where $\sigma > 0$ is a fixed parameter and $Z \sim N(0,1)$ is a standard normal random variable independent of X and Y. They have distribution functions F_σ and G_σ with densities

$$p_\sigma(x) = \int_{-\infty}^{\infty} \varphi_\sigma(x-y)\,dF(y) = -\frac{1}{\sigma^2}\int_{-\infty}^{\infty} F(x-y)\,y\,\varphi_\sigma(y)\,dy,$$
$$q_\sigma(x) = \int_{-\infty}^{\infty} \varphi_\sigma(x-y)\,dG(y) = -\frac{1}{\sigma^2}\int_{-\infty}^{\infty} G(x-y)\,y\,\varphi_\sigma(y)\,dy.$$

These identities easily imply:

Proposition B.1. (a) $\sup_x |p_\sigma(x) - q_\sigma(x)| \le \frac{1}{\sigma}\,\|F-G\|$; (b) $\|F_\sigma - G_\sigma\|_{\mathrm{TV}} \le \frac{1}{\sigma}\,W_1(F,G)$.

Thus, if F is close to G in a weak sense, then the regularized distributions will be closed in a stronger sense, at least when σ is not very small. One may replace W_1 in part (b) and the Kolmogorov distance in part (a) with other metrics:

Proposition B.2. *If* $\int_{-\infty}^{\infty} x^2\,dF(x) \le B^2$ *and* $\int_{-\infty}^{\infty} x^2\,dG(x) \le B^2$ $(B \ge 0)$, *then*

(a) $\|F_\sigma - G_\sigma\|_{\mathrm{TV}} \le \frac{2}{\sigma}\,\big[L(F,G) + 2B\,L(F,G)^{1/2}\big]$;
(b) $\|F_\sigma - G_\sigma\|_{\mathrm{TV}} \le \frac{4B}{\sigma}\,\|F-G\|^{1/2}$.

Proposition B.3. $\sup_x |p_\sigma(x) - q_\sigma(x)| \le \frac{L(F,G)}{\sigma}\,(1 + \frac{1}{2\sigma})$.

Appendix C. Special bounds for entropic distance to the normal

Let X be a random variable with mean zero and variance $\mathrm{Var}(X) = v^2$ $(v > 0)$ and with a bounded density p. In this section we formulate bounds for the entropic distance $D(X)$ in terms of the quadratic tail function $\delta_X(T) = \int_{\{|x| \ge T\}} x^2\,p(x)\,dx$ and another quantity, which is directly responsible for the closeness to the normal law,

$$\Delta(X) = \operatorname{ess\,sup}_x\,(p(x) - \varphi_v(x)).$$

As before, φ_v stands for the density of a normal random variable $Z \sim N(0, v^2)$, and we write φ in the standard case $v = 1$. The functional $\Delta(X)$ is homogeneous with respect to X with power of homogeneity -1 in the sense that $\Delta(\lambda X) = \Delta(X)/\lambda (\lambda > 0)$. Hence, the functional $\sqrt{\mathrm{Var}(X)}\,\Delta(X)$ is invariant under rescaling of the coordinates. To relate $D(X)$ and $\Delta(X)$, write $p(x) \le \varphi_v(x) + \Delta \le \frac{1}{v\sqrt{2\pi}} + \Delta$, so $p(x) \cdot v\sqrt{2\pi} \le 1 + \Delta\,v\sqrt{2\pi}$. This gives:

Proposition C.1. *Let X be a random variable with mean zero and variance* $\mathrm{Var}(X) = v^2$ $(v > 0)$*, having a bounded density. Then*

$$D(X) \le \log\Big(1 + v\Delta(X)\sqrt{2\pi}\Big) + \frac{1}{2}.$$

This estimate cannot, however, be used to see that X is almost normal. So, we need to refine Proposition C.1 for the case, where $\Delta(X)$ is small.

Proposition C.2. *Let X be a random variable with mean zero and variance* $\mathrm{Var}(X) = 1$*, having a bounded density. For all* $T \ge 0$*,*

$$D(X) \le \Delta(X)\left[\sqrt{2\pi} + 2T + 2T\ \log\big(1 + \Delta(X)\sqrt{2\pi}\, e^{T^2/2}\big)\right] + \frac{1}{2}\,\delta_X(T).$$

Hence, if $\Delta(X)$ is small and T is large, but not much, the right-hand side can be made small. When $\Delta(X) \le \frac{1}{2}$, one may take $T = \sqrt{2\log(1/\Delta(X))}$ which leads to the estimate

$$D(X) \le C\,\Delta(X)\sqrt{\log(1/\Delta(X))} + \frac{1}{2}\,\delta_X(T),$$

where C is absolute constant. If X satisfies the tail condition $\mathbf{P}\{|X| \ge t\} \le Ae^{-t^2/2}$ $(t > 0)$, we have $\delta_X(T) \le cA\,(1 + T^2)\,e^{-T^2/2}$ and then $D(X) \le C_A\,\Delta(X)\log\frac{1}{\Delta(X)}$.

References

[1] S.G. Bobkov, G.P. Chistyakov, F. Götze, Entropic instability of Cramer's characterization of the normal law, in: Selected works of Willem van Zwet, in: Sel. Works Probab. Stat., Springer, New York, 2012, pp. 231–242.

[2] S.G. Bobkov, G.P. Chistyakov, F. Götze, Stability problems in Cramer-type characterization in case of i.i.d. summands, Theory Probab. Appl. 57 (4) (2012) 701–723.

[3] S.G. Bobkov, G.P. Chistyakov, F. Götze, Regularized distributions and entropic stability of Cramer's characterization of the normal law. Extended version — arXiv:1504.02961v1 [math.PR] 12 Apr 2015.

[4] S.G. Bobkov, N. Gozlan, C. Roberto, P.-M. Samson, Bounds on the deficit in the logarithmic Sobolev inequality, J. Funct. Anal. 267 (11) (2014) 4110–4138.

[5] E.A. Carlen, A. Soffer, Entropy production by block variable summation and central limit theorems, Comm. Math. Phys. 140 (2) (1991) 339–371.

[6] G.P. Chistyakov, The sharpness of the estimates in theorems on the stability of decompositions of a normal distribution and a Poisson distribution, Teor. Funkcii Funkcional. Anal. i Prilozhen. Vyp. 26 (1976) 119–128. iii (in Russian).

[7] G.P. Chistyakov, L.B. Golinskii, Order-sharp estimates for the stability of decompositions of the normal distribution in the Levy metric, J. Math. Sci. 72 (1) (1994) 2848–2871 (in Russian), Translated in; Stability Problems for Stochastic Models, Vsesoyuz. Nauchno-Issled. Inst. Sistem. Issled., Moscow, 1991, pp. 16–40. (in Russian).

[8] H. Cramér, Ueber eine eigenschaft der normalen verteilungsfunktion, Math. Z. Bd. 41 (1936) 405–414.

[9] A. Dembo, T.M. Cover, J.A. Thomas, Information-theoretic inequalities, IEEE Trans. Inform. Theory 37 (6) (1991) 1501–1518.

[10] A. Figalli, F. Maggi, A. Pratelli, A refined Brunn–Minkowski inequality for convex sets, Ann. Inst. H. Poincaré Anal. Non Linéaire 26 (6) (2009) 2511–2519.

[11] A. Figalli, F. Maggi, A. Pratelli, Sharp stability theorems for the anisotropic Sobolev and log-Sobolev inequalities on functions of bounded variation, Adv. Math. 242 (2013) 80–101.

[12] M. Ledoux, Concentration of measure and logarithmic Sobolev inequalities, in: Seminaire de Probabilites, XXXIII, in: Lecture Notes in Math., vol. 1709, Springer, Berlin, 1999, pp. 120–216.

[13] Yu.V. Linnik, A remark on Cramer's theorem on the decomposition of the normal law, Teor. Veroyatn. Primen. 1 (1956) 479–480 (in Russian).

[14] Yu.V. Linnik, General theorems on the factorization of infinitely divisible laws. III. Sufficient conditions (countable bounded Poisson spectrum; unbounded spectrum; "stability"), Theor. Probab. Appl. 4 (1959) 142–163.

[15] Yu.V. Linnik, I.V. Ostrovskii, Decompositions of random variables and random vectors, Izd. Nauka, Moscow, 1972, p. 479 (in Russian). Translations of Mathematical Monographs, vol. 48, American Math. Society, Providence (translated from the Russian) R. I., 1977, ix+380 pp.
[16] S.G. Maloshevskii, Unimprovability of N. A. Sapogov's result in the stability problem of H. Cramer's theorem, Teor. Verojatn. Primen. 13 (1968) 522–525 (in Russian).
[17] H.P. McKean Jr., Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas, Arch. Ration. Mech. Anal. 21 (1966) 343–367.
[18] N.A. Sapogov, The stability problem for a theorem of Cramér, Izvestiya Akad. Nauk SSSR. Ser. Mat. 15 (1951) 205–218 (in Russian).
[19] N.A. Sapogov, The problem of stability for a theorem of Cramér, Vestn. Leningr. Univ. 10 (11) (1955) 61–64 (in Russian).
[20] A. Segal, Remark on stability of Brunn–Minkowski and isoperimetric inequalities for convex bodies, in: Geometric Aspects of Functional Analysis, in: Lecture Notes in Mathematics, vol. 2050, 2012, pp. 381–391.
[21] V.V. Senatov, Refinement of estimates of stability for a theorem of H. Cramér, in: Continuity and Stability in Problems of Probability Theory and Mathematical Statistics, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 61 (1976) 125–134 (in Russian).
[22] I.S. Shiganov, Some estimates connected with the stability of H. Cramer's theorem, in: Studies in Mathematical Statistics, vol. 3, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 87 (1979) 187–195 (in Russian).
[23] I.S. Shiganov, On Stability Estimates of Cramer's Theorem. Stability Problems for Stochastic Models, Varna, 1985, in: Lecture Notes in Math., vol. 1233, Springer, Berlin, 1987, pp. 178–181.
[24] V.M. Zolotarev, On the problem of the stability of the decomposition of the normal law into components, Teor. Verojatn. Primen. 13 (1968) 738–742 (in Russian).
[25] V.M. Zolotarev, Estimates of the difference between distributions in the Levy metric, in: Collection of Articles Dedicated to Academician Ivan Matveevich Vinogradov on His Eightieth Birthday, I. Trudy Mat. Inst. Steklov. 112 (1971) 224–231 388 (in Russian).