Chapter 15
Entropic instability of Cramer's characterization of the normal law

S.G. Bobkov and G.P. Chistyakov and F. Götze

Abstract We establish instability of the characterization of the normal law in Cramer's theorem with respect to the total variation norm and the entropic distance. Two constructions of counter-examples are provided.

15.1 Introduction

A well-known theorem of Cramer (1936, [Cr]) indicates that, if the sum $X + Y$ of two independent random variables X and Y has a normal distribution, then necessarily both X and Y are normal. Soon after Cramér had proved his theorem (which answered a question raised by P. Lévy in 1931), P. Lévy established stability of this characterization property of normal distributions. In a qualitative form it states that, for independent random variables X and Y,

> if $X + Y$ is nearly normal then both X and Y are nearly normal.

Here "nearly" is understood in the sense of the topology of weak convergence of probability distributions on the real line. For example, with respect to the Lévy distance, Lévy's theorem is formulated as follows. Given $\varepsilon > 0$ and distribution functions F_1, F_2,

Sergey G. Bobkov
School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455, USA
e-mail: bobkov@math.umn.edu

Gennadiy P. Chistyakov
Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany
e-mail: chistyak@math.uni-bielefeld.de

Friedrich Götze
Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany
e-mail: goetze@math.uni-bielefeld.de

$$L(F_1 * F_2, \Phi) < \varepsilon \quad \Rightarrow \quad L(F_1, \Phi_{a_1, \sigma_1}) < \delta_\varepsilon, \; L(F_2, \Phi_{a_2, \sigma_2}) < \delta_\varepsilon,$$

for some $a_1, a_2 \in \mathbf{R}$ and $\sigma_1, \sigma_2 > 0$, where δ_ε only depends on ε, and in a such way that $\delta_\varepsilon \to 0$, as $\varepsilon \to 0$.

Here $\Phi_{a,\sigma}$ stands for the distribution functions of the normal law $N(a, \sigma^2)$ with mean a and variance σ^2, and we omit indices in the standard case $a = 0$, $\sigma = 1$. As usual, $F_1 * F_2$ denotes the convolution of the distribution functions.

In 1950s Linnik [L2] extended this result to arbitrary probability distributions on the real line: If the convolution $F_1 * F_2$ is close to F, then both F_1 and F_2 have to be close to the class of all components of F. Linnik noted as well that Cramér's theorem may be viewed as a particular case of Darmois–Skitovich's theorem on the independence of independent linear statistics (cf. [L1]).

Another important issue which attracted many researchers is the problem of quantitative versions of the stability property of the normal law. This problem has been studied for a long time, starting with results by Sapogov in the 1950s [S1-2] (who considered the Kolmogorov distance and was apparently unaware of the work of P. Lévy) and ending with results by Chistyakov and Golinskii [C-G] in the 1990s, who found the correct asymptotics of the best possible error function $\varepsilon \to \delta_\varepsilon$ for the Lévy distance. See also [Z], [Se].

In this note we address the following natural question in connection with Lévy's theorem. Given independent random variables X and Y, assume that the distribution of $X + Y$ is known to be nearly normal in a stronger sense. What does this imply for X and Y in terms of closeness to the normal? When saying "stronger", we mean classical distances between distributions such as the total variation norm $\|F - G\|_{\mathrm{TV}}$, or the entropic distance $D(X)$ from a given distribution F of X to the associated normal law. Thus, we wonder whether or not X and Y need to be nearly normal with respect to these distances. In case of the entropic distance, this question was raised in the mid 1960's by Kac and McKean ([MC], pp. 365–366; cf. also [C-S] for some related aspects of the problem).

As it turns out, in general the answer is negative in both cases.

Theorem 1. *For any $\varepsilon > 0$, there exist independent random variables X and Y with absolutely continuous symmetric distributions F_1, F_2, and with $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$, such that*

a) $\|F_1 * F_2 - \Phi * \Phi\|_{\mathrm{TV}} < \varepsilon$;

b) $\|F_1 - \Phi_{a,\sigma}\|_{\mathrm{TV}} > c$ *and* $\|F_2 - \Phi_{a,\sigma}\|_{\mathrm{TV}} > c$, *for all* $a \in \mathbf{R}$ *and* $\sigma > 0$,

where $c > 0$ denotes an absolute constant.

As we will see, Theorem 1 holds for any number $c \in (0, 1/2)$.

The statement of the theorem may be strengthened in terms of the entropic distance. 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, what is classical, it 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 $\varphi_{a,\sigma}$ stands for the density of the normal law $N(a, \sigma^2)$ with parameters $a = \mathbb{E}X$, $\sigma^2 = \mathrm{Var}(X)$. Alternatively, it may be described as the shortest distance from the distribution F of X to the family of all normal laws on the line in the sense of the Kullback-Leibler distance.

Similarly to the total variation, the quantity $D(X)$ is homogeneous of order zero with respect to X, that is, $D(\lambda X) = D(X)$, for all $\lambda > 0$. In particular, it does not depend on the variance of X. The two distances are related by virtue of the Pinsker-Csiszár-Kullback inequality ([P], [Cs], [K]), which gives

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

In this sense the entropic distance is stronger than the total variation. Therefore, one may wonder whether or not the stability property in Cramer's theorem still holds when replacing the Lévy distance with the entropic distance. If so, this could also be viewed as the inverse to the concavity of the entropy functional (or to the so-called entropy power inequality, cf. [D-C-T]), which implies that

$$D(X + Y) \leq \frac{D(X) + D(Y)}{2},$$

whenever X and Y are independent and have equal variances.

It turns out however, this is not the case.

Theorem 2. *For any $\varepsilon > 0$, there exist independent random variables X and Y with absolutely continuous symmetric distributions F_1, F_2, and with $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$, such that*

a) $D(X + Y) < \varepsilon$;

b) $\|F_1 - \Phi_{a,\sigma}\|_{\mathrm{TV}} > c$ *and* $\|F_2 - \Phi_{a,\sigma}\|_{\mathrm{TV}} > c$, *for all $a \in \mathbf{R}$ and $\sigma > 0$,*

where $c > 0$ denotes an absolute constant. In particular, both $D(X)$ and $D(Y)$ are separated from zero.

In the next section we describe how such random variables may be constructed. In fact, our (counter-)examples for Theorem 1 still work for Theorem 2. We consider two constructions. The first one explicitly specifies densities for X and Y, while the

other one deals with their distribution functions, which are explicitly provided, as well.

In Section 3 we show that the distributions of X and Y are separated from the normal law, thus proving claim b) of Theorem 1. Finally, in Section 4 we provide computations for the convolutions, which will justify claim a) of Theorem 1.

15.2 Constructions of examples

In this section we describe two types of the construction of random variables.

We use the standard notations

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \quad \Phi(x) = \int_{-\infty}^{x} \varphi(y)\, dy \qquad (x \in \mathbf{R})$$

for the density and the distribution function of the standard normal law.

Construction I (by an explicit formula for densities).
Given $T > 0$, let X_T be a random variable with density function

$$p_T(x) = c_T \sin^2(Tx)\, \varphi(x), \qquad x \in \mathbf{R},$$

where $c_T = 2/(1 - e^{-2T^2})$ is the normalizing constant. Introduce a further random variable, X_{2T}, independent of X_T, with density p_{2T}.

Clearly, X_T has a symmetric distribution with

$$\mathbb{E}X_T^2 = \frac{c_T}{2}\left(1 - (1 - 4T^2)e^{-2T^2}\right) \to 1, \quad \text{as } T \to +\infty.$$

Based on this choices, in the proof of Theorems 1–2 we consider

$$X = \frac{X_T}{\sqrt{\mathbb{E}X_T^2}}, \quad Y = \frac{X_{2T}}{\sqrt{\mathbb{E}X_{2T}^2}}$$

for large values of T.

Note that we may rewrite our densities as

$$p_T(x) = \frac{c_T}{2}\left(\varphi(x) - \cos(2Tx)\,\varphi(x)\right).$$

As another variant one may also consider densities of the form

$$p(x) = \varphi(x) + \sin(Tx)\,\varphi(x),$$

which are somewhat simpler. However, they are not symmetric about the origin.

Construction II (by an explicit formula for distribution functions).
Given $T > 0$, let X_T be a random variable with the distribution function

$$F_T(x) = \Phi(x) + \frac{1}{2T}\, \sin(Tx)\, \varphi(x)\, 1_{\{|x| < T\}}.$$

Their densities are given by

$$p_T(x) = \varphi(x) + \frac{1}{2}\left(\cos(Tx) - \frac{x}{T}\sin(Tx)\right)\varphi(x)\, 1_{\{|x| < T\}}.$$

Clearly, $p_T(x) > 0$ everywhere (perhaps except for $|x| = T$), so F_T is increasing.
Since also $F_T(-\infty) = 0$, $F_T(+\infty) = 1$, F_T is indeed a distribution function. Note that
p_T is even, so the distribution of X_T is symmetric about the origin.

Again introduce a second independent random variable X_{2T} with the distribution
function F_{2T}.

To see that $\mathrm{Var}(X_T) \to 1$, as $T \to +\infty$, we may apply well-known identities which
can be obtained by the successive differentiation of the identity $\int_{-\infty}^{+\infty} \cos(Tx)\, \varphi(x)\, dx = e^{-T^2/2}$ with respect to the variable T:

1) $\int_{-\infty}^{+\infty} x \sin(Tx)\, \varphi(x)\, dx = T\, e^{-T^2/2}$,

2) $\int_{-\infty}^{+\infty} x^2 \cos(Tx)\, \varphi(x)\, dx = (1 - T^2)\, e^{-T^2/2}$,

3) $\int_{-\infty}^{+\infty} x^3 \sin(Tx)\, \varphi(x)\, dx = (3T - T^3)\, e^{-T^2/2}$.

Write

$$\mathbb{E}X_T^2 = 1 + \frac{1}{2}\int_{-T}^{T} x^2\left(\cos(Tx) - \frac{x}{T}\sin(Tx)\right)\varphi(x)\, dx.$$

By 2)–3), extending integration to the whole line, we get that

$$\mathbb{E}X_T^2 = 1 - e^{-T^2/2} - \frac{1}{2}\int_{\{|x| > T\}} x^2\left(\cos(Tx) - \frac{x}{T}\sin(Tx)\right)\varphi(x)\, dx.$$

Clearly, the last integral tends to zero.

Based on this choices, for the proof of Theorems 1–2 one may similarly take

$$X = \frac{X_T}{\sqrt{\mathbb{E}X_T^2}}, \quad Y = \frac{X_{2T}}{\sqrt{\mathbb{E}X_{2T}^2}}$$

for large values of T.

Although seemingly more complicated, the second construction is more con-
venient, when measuring the distance to the normal for metrics, such as Lévy
and Kantorovich-Rubinshtein, which explicitly involve distribution functions (rather
than densities).

15.3 Separation from the normal

The distributions F_T of X_T, constructed in the previous section, are close to the standard normal in the sense of the topology of weak convergence. To see this, let us look at the characteristic functions for the distributions from Construction I:

$$
\begin{aligned}
f_T(t) = \mathbb{E}\,e^{itX_T} &= \int_{-\infty}^{+\infty} e^{itx} p_T(x)\,dx \\
&= \frac{c_T}{2} \int_{-\infty}^{+\infty} \cos(tx)\,(1 - \cos(2Tx))\,\varphi(x)\,dx \\
&= \frac{c_T}{2} \int_{-\infty}^{+\infty} \left(\cos(tx) - \frac{\cos((t+2T)x) + \cos((t-2T)x)}{2} \right) \varphi(x)\,dx \\
&= \frac{c_T}{2} \left(e^{-t^2/2} - \frac{e^{-(t+2T)^2/2} + e^{-(t-2T)^2/2}}{2} \right).
\end{aligned}
$$

Hence, for any fixed real t,

$$
f_T(t) = \frac{1}{1 - e^{-2T^2/2}} \left(e^{-t^2/2} - \frac{e^{-(t+2T)^2/2} + e^{-(t-2T)^2/2}}{2} \right) \to e^{-t^2/2},
$$

and thus weakly in distribution

$$
X_T \Rightarrow N(0,1), \quad \text{as} \quad T \to +\infty.
$$

By a compactness argument, it is easy to see that $\rho(F_T, \Phi) \to 0$, for any metric metrizing the weak convergence in the space of all probability distributions on the line. If the second moments of distributions are known to be bounded, one may use, for example, the Kantorovich-Rubinshtein distance, which in our case is given by

$$
W_1(F_T, \Phi) = \int_{-\infty}^{+\infty} |F_T(x) - \Phi(x)|\,dx.
$$

By the very definition of the distributions from Construction II, we obtain immediately that $W_1(F_T, \Phi) < \frac{1}{2T}$.

As a consequence, the normalized random variables X and Y are also close to the standard normal law for the Kantorovich-Rubinshtein metric.

On the other hand, let us look at the total variation distance. One may apply the general elementary estimate

$$
\sup_{t \in \mathbf{R}} |f(t) - g(t)| \leq \|F - G\|_{\mathrm{TV}},
$$

holding for arbitrary probability distributions F and G on the real line with characteristic functions f and g, respectively. In particular, for the distributions from the first construction (choosing $t = 2T$), we have

$$\|F_T - \Phi\|_{\mathrm{TV}} \geq \sup_{t \in \mathbf{R}} |f_T(t) - e^{-t^2/2}| \geq |f_T(2T) - e^{-2T^2}| \to \frac{1}{2},$$

as $T \to +\infty$. Hence,

$$\liminf_{T \to +\infty} \|F_T - \Phi\|_{\mathrm{TV}} \geq \frac{1}{2}.$$

This observation can be strengthened by considering the shortest total variation distance from F_T to the class of all normal laws on the line.

Lemma 1. *We have*

$$\liminf_{T \to +\infty} \inf_{a,\sigma} \|F_T - \Phi_{a,\sigma}\|_{\mathrm{TV}} \geq \frac{1}{2}.$$

Proof. As was discussed above, we may use the bounds

$$\|F_T - \Phi_{a,\sigma}\|_{\mathrm{TV}} \geq \sup_{t \in \mathbf{R}} \left| f_T(t) - e^{iat - \sigma^2 t^2/2} \right|$$

$$\geq \sup_{t \in \mathbf{R}} \left| |f_T(t)| - e^{-\sigma^2 t^2/2} \right|.$$

It follows from the formula for $f_T(t)$ that uniformly over all $t \geq 0$,

$$\left| f_T(t) - e^{-\sigma^2 t^2/2} \right| \geq \left| \left| e^{-t^2/2} - \frac{1}{2} e^{-(t-2T)^2/2} \right| - e^{-\sigma^2 t^2/2} \right| - o(T),$$

as $T \to +\infty$, so

$$\liminf_{T \to +\infty} \inf_{a,\sigma} \|F_T - \Phi_{a,\sigma}\|_{\mathrm{TV}} \geq \liminf_{T \to +\infty} \inf_{\sigma>0} \sup_{t>0} \left| \left| e^{-t^2/2} - \frac{1}{2} e^{-(t-2T)^2/2} \right| - e^{-\sigma^2 t^2/2} \right|.$$

Here and in the sequel, $o(T)$ denotes a quantity which tends to zero, as $T \to +\infty$, uniformly over all t from the indicated range.

To estimate the supremum on the right-hand side uniformly over all $\sigma > 0$, fix a (large) number N. In case $\sigma \geq N/T$, choose $t = 2T$, which gives

$$\left| \left| e^{-t^2/2} - \frac{1}{2} e^{-(t-2T)^2/2} \right| - e^{-\sigma^2 t^2/2} \right| = \frac{1}{2} - e^{-2\sigma^2 T^2} + o(T) \geq \frac{1}{2} - e^{-2N^2} + o(T).$$

In case $\sigma < N/T$, choose $t = 2\sqrt{T}$, which gives

$$\left| \left| e^{-t^2/2} - \frac{1}{2} e^{-(t-2T)^2/2} \right| - e^{-\sigma^2 t^2/2} \right| = e^{-2\sigma^2 T} + o(T) \geq e^{-2N^2/T} + o(T),$$

where the right-hand side tends to 1, as $T \to +\infty$. Altogether this yields

$$\liminf_{T \to +\infty} \inf_{a,\sigma} \|F_T - \Phi_{a,\sigma}\|_{\mathrm{TV}} \geq \frac{1}{2} - e^{-2N^2}.$$

Since the left-hand side does not depend on N, we may let $N \to +\infty$, and the lemma follows.

As we mentioned in the previous section, the random variables X and Y in Theorems 1–2 are obtained from X_T and X_{2T} by normalizing, so that $\mathrm{Var}(X) = \mathrm{Var}(Y)$. Since the total variation norm is invariant under rescaling of the coordinates, Lemma 1 also implies that,

$$\liminf_{T \to +\infty} \inf_{a,\sigma} \|F - \Phi_{a,\sigma}\|_{\mathrm{TV}} \geq \frac{1}{2}, \qquad \liminf_{T \to +\infty} \inf_{a,\sigma} \|G - \Phi_{a,\sigma}\|_{\mathrm{TV}} \geq \frac{1}{2},$$

where F and G denote distributions of X and Y (which also depend on T).

Recalling also Pinsker-Csiszár-Kullback's inequality, we may conclude the property $b)$ in these theorems.

Conclusion 1. For random variables X and Y from Construction I, we have

$$\|F - \Phi_{a,\sigma}\|_{\mathrm{TV}} > c, \qquad \|G - \Phi_{a,\sigma}\|_{\mathrm{TV}} > c,$$

for all T large enough, where c is any prescribed number in $(0, 1/2)$. In particular, $D(X) > c^2/4$ and $D(Y) > c^2/4$.

A similar approach may be used to study the distributions F_T from the second construction. The corresponding characteristic functions are given by

$$f_T(t) = e^{-t^2/2} + \frac{1}{2} \int_{-T}^{T} e^{itx} \left(\cos(Tx) - \frac{x}{T} \sin(Tx) \right) \varphi(x)\, dx$$

$$= e^{-t^2/2} + \frac{e^{-(t+T)^2/2} + e^{-(t-T)^2/2}}{4}$$

$$\quad - \frac{1}{2} \int_{|x|>T} e^{itx} \cos(Tx)\, \varphi(x)\, dx - \frac{1}{2T} \int_{-T}^{T} e^{itx} x \sin(Tx)\, \varphi(x)\, dx.$$

Clearly, the first integral is bounded in absolute value by $2(1 - \Phi(T)) < e^{-T^2/2}$, while the absolute value of the second integral is smaller than $\int |x|\, \varphi(x)\, dx < 1$. Hence, uniformly over all $t \in \mathbf{R}$

$$f_T(t) = e^{-t^2/2} + \frac{e^{-(t+T)^2/2} + e^{-(t-T)^2/2}}{4} + o(T),$$

as $T \to +\infty$. Next one can repeat the line of arguments from the proof of Lemma 1.

Conclusion 2. For the random variables X and Y of Construction II, Conclusion 1 holds with level $1/4$ replacing $1/2$ (for constants c).

15.4 Convolutions of distributions from Construction I

Write the density of random variables X_T from Construction I in the form

$$p_T(x) = \frac{c_T}{2} \left(\varphi(x) - \cos(2Tx)\,\varphi(x) \right),$$

where $c_T = 2/(1 - e^{-2T^2})$ is the normalizing constant. Note that $\frac{c_T}{2} \to 1$, as $T \to +\infty$.

Instead of the sum $X + Y$ (which is a bit more complicated), we consider the sum $X_T + X_{2T}$ of two independent random variables, assuming that X_T has density p_T and X_{2T} has density p_{2T}. The density of this sum represents the convolution $p_T * p_{2T}$.

In analogy with notations for distribution functions, for integrable functions $p(x)$ and $q(x)$ we write $(p * q)(x) = p(x) * q(x) = \int_{-\infty}^{+\infty} p(x - y) q(y)\,dy$.

To simplify the computations, introduce

$$q_T(x) = \varphi(x) - \cos(2Tx)\,\varphi(x)$$

and write

$$\begin{aligned}
(q_T * q_{2T})(x) - (\varphi * \varphi)(x) = &- \varphi(x) * [(\cos(2Tx) + \cos(4Tx))\,\varphi(x)] \\
&+ [\cos(2Tx)\,\varphi(x)] * [\cos(4Tx))\,\varphi(x)].
\end{aligned}$$

Note that

$$(\varphi * \varphi)(x) = \frac{1}{2\sqrt{\pi}}\, e^{-x^2/4}.$$

To compute convolutions, we need one simple relation. Given a complex variable α, consider the integral

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

Changing the variable $y = \frac{x}{2} - \frac{t}{\sqrt{2}}$, we obtain $\frac{(x-y)^2 + y^2}{2} = \frac{x^2}{4} + \frac{t^2}{2}$, and the integral becomes

$$\frac{1}{\sqrt{\pi}}\,\frac{1}{\sqrt{2}}\, e^{-x^2/4}\, e^{\alpha x/2} \int_{-\infty}^{+\infty} e^{-\alpha t/\sqrt{2}}\, e^{-t^2/2}\,dt = e^{-x^2/4}\, e^{\alpha x/2 + \alpha^2/4}.$$

Therefore,

$$\int_{-\infty}^{+\infty} \varphi(x - y)\,\varphi(y)\, e^{\alpha y}\,dy = (\varphi * \varphi)(x)\, e^{\alpha x/2 + \alpha^2/4}.$$

Taking $\alpha = iT$, we get

$$A_T \equiv \int_{-\infty}^{+\infty} \varphi(x - y)\,\varphi(y)\, \cos(Ty)\,dy = (\varphi * \varphi)(x)\, e^{-T^2/4} \cos(Tx/2),$$

$$B_T \equiv \int_{-\infty}^{+\infty} \varphi(x-y)\,\varphi(y)\,\sin(Ty)\,dy = (\varphi * \varphi)(x)\,e^{-T^2/4}\,\sin(Tx/2).$$

Hence, the convolution $\varphi(x) * [(\cos(2Tx) + \cos(4Tx))\,\varphi(x)]$ is given by

$$\int_{-\infty}^{+\infty} \varphi(x-y)\,\varphi(y)\,(\cos(2Ty)+\cos(4Ty))\,dy = A_{2T} + A_{4T}.$$

Similarly, the convolution $[\cos(2Tx)\,\varphi(x)] * [\cos(4Tx))\,\varphi(x)]$ is

$$\int_{-\infty}^{+\infty} \varphi(x-y)\,\varphi(y)\,\cos(2T(x-y))\,\cos(4Ty))\,dy$$
$$= \int_{-\infty}^{+\infty} \varphi(x-y)\,\varphi(y)\,\frac{\cos(2Tx-6Ty)+\cos(2Tx+2Ty)}{2}\,dy$$
$$= \tfrac{1}{2}\,\left(A_{6T}\cos(2Tx)+A_{2T}\cos(2Tx)+B_{6T}\sin(2Tx)-B_{2T}\sin(2Tx)\right).$$

Collecting the two convolutions together, we obtain for $(q_T * q_{2T})(x)$ the representation

$$(\varphi * \varphi)(x) - A_{2T} - A_{4T} + \frac{A_{6T}+A_{2T}}{2}\cos(2Tx) + \frac{B_{6T}-B_{2T}}{2}\sin(2Tx).$$

Now, using the obvious bound $|A_{2T}| \leq (\varphi * \varphi)(x)\,e^{-T^2}$ and similarly for B_{2T}, we arrive at

$$\left| \frac{(q_T * q_{2T})(x)}{(\varphi * \varphi)(x)} - 1 \right| \leq 4e^{-T^2}.$$

But $p_T * p_{2T} = (1+\varepsilon_T)\,q_T * q_{2T}$, where $\varepsilon_T = \tfrac{1}{4}c_T c_{2T} - 1 \to 0$, as $T \to +\infty$, and moreover $|\varepsilon_T| \leq Ce^{-2T^2}$, whenever $T \geq 1$. Hence, we get:

Lemma 2. *For all $T \geq 1$ and $x \in \mathbf{R}$,*

$$\left| \frac{(p_T * p_{2T})(x)}{(\varphi * \varphi)(x)} - 1 \right| \leq Ce^{-T^2}$$

with some absolute constant C.

This estimate is quite sufficient to see that

$$\|F_T * F_{2T} - \Phi * \Phi\|_{\mathrm{TV}} \leq Ce^{-T^2}$$

and also for the Kullback-Leibler's distance

$$D(X_T + X_{2T} \| Z) = \int_{-\infty}^{+\infty} (p_T * p_{2T})(x) \log \frac{(p_T * p_{2T})(x)}{(\varphi * \varphi)(x)}\,dx \to 0,$$

as $T \to +\infty$, where $Z \sim N(0,2)$. So the (closest) entropic distance to the normal

$$D(X_T + X_{2T}) \to 0.$$

A similar property, $D(X + Y) \to 0$, as $T \to +\infty$, also holds for normalized random variables, since $\text{Var}(X_T) \to 1$, although this conclusion requires a certain justification. What is needed is the property

$$D(\alpha_T X_T + X_{2T}) \to 0$$

where $\alpha_T \to 1$. This may be done, for example, by a slight modification of the arguments used in the proof of Lemma 2. With this in mind Theorems 1–2 are proved.

We leave it to the reader to check that the same conclusion is true for probability distributions from Construction II.

References

[C-S] Carlen, E. A., Soffer, A. Entropy production by block variable summation and central limit theorems. Comm. Math. Phys. 140 (1991), no. 2, 339–371.

[C-G] Chistyakov, G. P., Golinskii, L. B. Order-sharp estimates for the stability of decompositions of the normal distribution in the Levy metric. (Russian) Translated in J. Math. Sci. 72 (1994), no. 1, 2848–2871. Stability problems for stochastic models (Russian) (Moscow, 1991), 16–40, Vsesoyuz. Nauchno-Issled. Inst. Sistem. Issled., Moscow, 1991.

[Cr] Cramér, H. Ueber eine Eigenschaft der Normalen Verteilungsfunktion. Math. Zeitschrift, Bd. 41 (1936), 405–414.

[Cs] Csiszár, I. Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar., 2 (1967), 299–318.

[D-C-T] Dembo, A., Cover, T. M., Thomas, J. A. Information-theoretic inequalities. IEEE Trans. Inform. Theory, 37 (1991), no. 6, 1501–1518.

[K] Kullback, S. A lower bound for discrimination in terms of variation. IEEE Trans. Inform. Theory, T-13, 1967, 126–127.

[L1] Linnik, Yu. V. A remark on Cramer's theorem on the decomposition of the normal law. (Russian) Teor. Veroyatnost. i Primenen. 1 (1956), 479–480.

[L2] Linnik, Yu. V. General theorems on the factorization of infinitely divisible laws. III. Sufficient conditions (countable bounded Poisson spectrum; unbounded spectrum; "stability"). Theor. Probability Appl. 4 (1959), 142–163.

[MK] McKean, H. P., Jr. Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas. Arch. Rational Mech. Anal. 21 (1966), 343–367.

[P] Pinsker, M. S. Information and information stability of random variables and processes. Translated and edited by Amiel Feinstein Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam, 1964, xii+243 pp.

[S1] Sapogov, N. A. The stability problem for a theorem of Cramér. (Russian) Izvestiya Akad. Nauk SSSR. Ser. Mat. 15 (1951), 205–218.

[S2] Sapogov, N. A. The problem of stability for a theorem of Cramér. (Russian) Vestnik Leningrad. Univ. 10 (1955), no. 11, 61–64.

[Se] Senatov, V. V. Refinement of estimates of stability for a theorem of H. Cramér. (Russian) Continuity and stability in problems of probability theory and mathematical statistics. Zap. Naucn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 61 (1976), 125–134.

[Z] Zolotarev, V. M. On the problem of the stability of the decomposition of the normal law into components. (Russian) Teor. Verojatnost. i Primenen., 13 (1968), 738–742.