Acta Applicandae Mathematicae **58:** 159–174, 1999.
© 1999 *Kluwer Academic Publishers. Printed in the Netherlands.*

Pre-limit Theorems and Their Applications

L. B. KLEBANOV*
St. Petersburg State University for Architecture and Civil Engineering, 2nd Krasnoarmeyskaya, 4, 198005 St. Petersburg, Russia. e-mail: lev@pdmi.ras.ru

S. T. RACHEV
Department of Economics, University of Karlsruhe, Postfach 6980, D-76128 Karlsruhe, Germany. e-mail: zari.rachev@wiwi.uni-karlsruhe.de

G. J. SZEKELY
Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio 43403, USA. e-mail: gabors@bgnet.bgsu.edu

(Received: 28 August 1999)

Abstract. There exists a considerable debate in the literature about the applicability of α-stable distributions as they appear in Lévy's central limit theorems. A serious drawback of Lévy's approach is that, in practice, one can never know whether the underlying distribution is heavy tailed, or just has a long but truncated tail. Limit theorems for stable laws are not robust with respect to truncation of the tail or with respect to any change from 'light' to 'heavy' tail, or conversely. In this talk we provide a new 'pre-limiting' approach that helps overcome this drawback of Lévy-type central limit theorems.

Mathematics Subject Classifications (1991): Primary: 60E07; Secondary: 60E05, 60F05, 90A09, 90A12.

Key words: prelimiting behavior, stable distribution, random stability.

1. Introduction and Statement of the Problem

Finitely many empirical observations can never justify any tail behavior, thus they cannot justify the applicability of classical limit theorems in probability theory. In this paper we attempt to show that instead of relying on limit theorems, one may use the so-called pre-limit theorems explained later. The applicability of our prelimit theorem relies not on the tail but on the 'central section' ('body') of the distributions and as a result, instead of a limiting behavior (when n, the number of i.i.d. observations tends to infinity), the pre-limit theorem should provide an approximation for distribution functions in case n is 'large' but not too 'large'.

Our pre-limiting approach seems to be more realistic for practical applications. We shall start with two examples.

* The research was supported by the German–Russian Grant 98–01–04070 and by 1999 Grant of RTBR.

EXAMPLE 1. *Pareto-stable laws.* More than a hundred years ago, Vilfredo Pareto [19] observed that the number of people in the population whose income exceeds a given level x can be satisfactorily approximated by $Cx^{-\alpha}$ for some C and $\alpha > 0$. (See [1, 3, 12] for more details.) Later, Mandelbrot [14, 15] argued that stable laws should provide the appropriate model for income distributions; after some statistical studies on income data he made two claims:

(i) the distribution of the size of income for different (but sufficiently long) time periods must be of the same type, in other words the distribution of the income follows a stable law (Lévy's stable law, see [2]),
(ii) the tails of the Gaussian law are too thin to describe the distribution of the income in typical situations, see [16, 17].

It is known that the variance of any non-Gaussian stable law is infinite, thus an essential condition for a non-Gaussian stable limit distribution for sums of random incomes is that the summands have 'heavy' tails in the sense that the variance of the summands must be infinite. On the other hand, it is obvious that the incomes are always bounded random variables (in view of the finiteness of all available money in the world, and the existence of a smallest monetary unit). Even if we assume that the support of the income distribution is infinite, there exists a considerable amount of empirical studies showing that the income distributions have Pareto tails with index α between 3 and 4, so the variance is finite, see [4]. Thus, in practice the underlying distribution *cannot* be heavy tailed. *Does this mean that we have to reject the Pareto-stable model?*

EXAMPLE 2. *Exponential decay.* One of the most popular examples for exponential distributions is the random time for radioactive decay. The exponential distribution is in the domain of attraction of the Gaussian law. In quantum physics it has been shown [8, 23, 20] that theoretically the radioactive decay is not exactly exponentially distributed. Recently, a new experimental evidence supported that conclusion (see [22]). But then one faces the following paradox.

To describe the model let $p(t)$ be the probability density that a physical system is in the initial state at moment $t \geqslant 0$. It is known (see, for example, [25, p. 42]) that $p(t) = |f(t)|^2$, where

$$f(t) = \int_0^\infty \omega(E) \exp(iEt) \, \mathrm{d}E,$$

and $\omega(E) \geqslant 0$ is the density of the energy of the disintegrating physical system. For a broad class of physical systems

$$\omega(E) = \frac{A}{(E - E_0)^2 + \Gamma^2}, \qquad E \geqslant 0$$

(see [25] and the references therein), where A is a normalizing constant, and E_0 and Γ are the mode and the measure of dissipation of the system energy (with respect to E_0). For typical nonstable physical systems, the ratio Γ/E_0 is very small (it is of order 10^{-15} or smaller). Therefore, we have that

$$f(t) = \mathrm{e}^{iE_0 t}\frac{A}{\Gamma}\int_{-E_0/\Gamma}^{\infty}\frac{\mathrm{e}^{i\Gamma t y}}{y^2+1}\,\mathrm{d}y$$

differs by a very small value (of magnitude 10^{-15}) from

$$f_1(t) = \mathrm{e}^{iE_0 t}\frac{A}{\Gamma}\int_{-\infty}^{\infty}\frac{\mathrm{e}^{i\Gamma t y}}{y^2+1}\,\mathrm{d}y = \pi\mathrm{e}^{iE_0 t}\frac{A}{\Gamma}\,\mathrm{e}^{-t\Gamma}, \quad t > 0.$$

That is, $p(t) = |f(t)|^2$ is approximately equal to $(\frac{\pi A}{\Gamma})^2\mathrm{e}^{-2t\Gamma}$, which gives (as an approximation) the classical exponential distribution model of decay. On the other hand, it is equally easy to find the asymptotic representation of $f(t)$ as $t \to \infty$. Namely,

$$\begin{aligned}\int_{-E_0/\Gamma}^{\infty}\frac{\mathrm{e}^{i\Gamma t y}}{y^2+1}\,\mathrm{d}y \;&=\; \int_{-\arctan(E_0/\Gamma)}^{\pi/2}\mathrm{e}^{i\Gamma t\tan z}\mathrm{d}z\\ &\sim\; -\frac{\cos^2(\arctan(E_0/\Gamma))}{it\Gamma}\,\mathrm{e}^{-itE_0}.\end{aligned}$$

Therefore,

$$f(t) \sim i\frac{A}{E_0^2+\Gamma^2}\frac{1}{t}, \quad \text{as } t \to \infty,$$

where

$$A = \frac{1}{\int_0^{\infty}\frac{\mathrm{d}E}{(E-E_0)^2+\Gamma^2}},$$

or

$$p(t) \sim \frac{A^2}{(E_0^2+\Gamma^2)^2}\frac{1}{t^2}, \quad \text{as } t \to \infty.$$

Therefore, $p(t)$ belongs to the domain of attraction of a stable law with index $\alpha = 1$. Thus, if T_j, $j \geqslant 1$, are i.i.d. r.v.'s describing the times of decay of a physical system, then the sum $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(T_j - c))$ does *not* tend to a Gaussian distribution for any centering constant c (as we could have expected under exponential decay), but diverges to infinity. *Does this mean that the exponential approximation cannot be used anymore?*

In the above examples we see that the problem of passing to limit distributions is 'ill-posed' in the sense that a small perturbation of the tail of the underlying distribution changes significantly the limit behavior of the normalized sum of r.v.'s.

We can see the same problem in a more general situation. Given i.i.d. r.v.'s X_j, $j \geqslant 1$, the limiting behavior of the normalized partial sums $S_n = n^{-1/\alpha}(X_1+\cdots+X_n)$ depends on the tail behavior of X. Both the proper normalization, $n^{-1/\alpha}$, in S_n and the corresponding limiting law are extremely sensitive to a tail truncation. We claim that in this sense the problem of limiting distributions for sums of i.i.d. r.v.'s is *ill-posed*. We shall propose a 'well-posed' version of the problem and provide a solution in the form of a *pre-limit theorem*.

Let us fix two positive constants c and γ, and consider the following semi-distance between the random variables X and Y:

$$d_{c,\gamma}(X, Y) = \sup_{|t|\geqslant c} \frac{|f_X(t) - f_Y(t)|}{|t|^\gamma}.$$

(Here and in what follows $F_Y(x)$ and $f_Y(t)$ stand for the cumulative distribution function (c.d.f.) and the characteristic function of X, respectively.)

Observe that in the case $c = 0$, $d_{c,\gamma}(X, Y)$ defines a well-known probability distance in the space of all r.v.'s for which $d_{0,\gamma}(X, Y)$ is finite, see [21, 24].

Next recall that Y is a strictly α-stable r.v. if for every positive integer n

$$Y_1 \stackrel{d}{=} U_n := \frac{Y_1 + \cdots + Y_n}{n^{1/\alpha}},$$

where $\stackrel{d}{=}$ stands for equality in distribution and Y_j, $j \geqslant 1$, are i.i.d. $Y_j \stackrel{d}{=} Y$, see [13, 25].

Let X, X_j, $j \geqslant 1$, be a sequence of i.i.d. r.v.'s such that $d_{0,\gamma}(X, Y)$ is finite for some strictly stable random variable Y. Suppose that Y, Y_j, $j \geqslant 1$, are i.i.d. strictly α-stable random variables, and $\gamma > \alpha$. Then

$$\begin{aligned} d_{0,\gamma}(S_n, Y) &= d_{0,\gamma}(S_n, U_n) \\ &= \sup_t \frac{|f_X^n(t/n^{1/\alpha}) - f_Y^n(t/n^{1/\alpha})|}{|t|^\gamma} \\ &\leqslant n \sup_t \frac{|f_X(t/n^{1/\alpha}) - f_Y(t/n^{1/\alpha})|}{|t|^\gamma} = \frac{1}{n^{\gamma/\alpha - 1}} d_{0,\gamma}(X, Y), \end{aligned}$$

see [25]. From this we can see that $d_{0,\gamma}(S_n, Y)$ tends to zero as n tends to infinity, that is, we have convergence (in $d_{0,\gamma}$) of the normalized sums of X_j to a strictly α-stable random variable Y provided that $d_{0,\gamma}(X, Y) < \infty$. However, *any* truncation of the tail of the distribution of X leads to $d_{0,\gamma}(X, Y) = \infty$. Our goal is to analyze the closeness of the sum S_n to a strictly α-stable random variable Y without the assumption on the finiteness of $d_{0,\gamma}(X, Y)$, restricting our assumptions to bounds in terms of $d_{c,\gamma}(X, Y)$ with $c > 0$. In this way we shall formulate a general type of a *central pre-limit theorem* with no assumption on the tail behavior of the underlying random variables. We shall illustrate our theorem by providing answers to the problems addressed in Examples 1 and 2.

2. Main Result

In our Central Pre-Limit Theorem we shall analyze the closeness of the sum S_n to a strictly α-stable r.v. Y in terms of the following Kolmogorov metric (see [11] and [21]): for any c.d.f.'s F and G,

$$k_h(F, G) := \sup_{x\in\mathbb{R}} |F * h(x) - G * h(x)|,$$

where $*$ stands for convolution, and the "smoothing" $h(x)$ is a fixed cdf with a bounded continuous density function, $\sup_x |h'(x)| \leqslant c(h) < \infty$. Metric k_h metrizes the weak convergence in the space of cdf's [11].

THEOREM 2.1 (Central Pre-Limit Theorem). *Let X, X_j, $j \geqslant 1$, be i.i.d. r.v.'s and $S_n = n^{-1/\alpha} \sum_{j=1}^n X_j$. Suppose that Y is a strictly α-stable r.v. Let $\gamma > \alpha$ and $\Delta > \delta$ be arbitrary given positive constants and let $n \leqslant (\Delta/\delta)^\alpha$ be an arbitrary positive integer. Then*

$$k_h(F_{S_n}, F_Y) \leqslant \inf_{a>0} \left(\sqrt{2\pi}\, \frac{d_{\delta,\gamma}(X, Y)(2a)^\gamma}{n^{\gamma/\alpha - 1}\gamma} + 2\frac{c(h)}{a} + 2\Delta a \right).$$

Remark 2.1. If $\Delta \to 0$ and, furthermore, $\Delta/\delta \to \infty$, then n can be chosen large enough so that the right-hand side of the above bound is sufficiently small, that is, we obtain the classical limit theorem for weak convergence to an α-stable law. This result, of course, includes the central limit theorem for weak distance.

Proof of the Theorem 2.1. For $\gamma > \alpha$,

$$\begin{aligned} d_{c,\gamma}(S_n, Y) &= d_{c,\gamma}(S_n, U_n) \\ &\leqslant n \sup_{|t| \geqslant c} \frac{|f_X(t/n^{1/\alpha}) - f_Y(t/n^{1/\alpha})|}{|t|^\gamma} = \frac{1}{n^{\gamma/\alpha - 1}} d_{\frac{c}{n^{1/\alpha}},\gamma}(X, Y). \end{aligned}$$

For any $\Delta > \delta$ and for all $n \leqslant (\Delta/\delta)^\alpha$, we have then

$$d_{\Delta,\gamma}(S_n, Y) \leqslant \frac{1}{n^{\gamma/\alpha - 1}} d_{\delta,\gamma}(X, Y).$$

The above relation can be rewritten in the form

$$\sup_{|t| \geqslant \Delta} \frac{|f_{S_n}(t) - f_Y(t)|}{|t|^\gamma} \leqslant \frac{1}{n^{\gamma/\alpha - 1}} d_{\delta,\gamma}(X, Y).$$

Denote by $I(t)$ the indicator function of the interval $[-\Delta, \Delta]$, then

$$\frac{1}{|t|} \big| (1 - I(t)) f_{S_n}(t) - (1 - I(t)) f_Y(t) \big| \leqslant \frac{|t|^{\gamma - 1}}{n^{\gamma/\alpha - 1}} d_{\delta,\gamma}(X, Y).$$

For any $a > 0$ define

$$\widetilde{V}_a(t) = \sqrt{\frac{\pi}{2}} \begin{cases} 1 & \text{for } |t| < a, \\ \dfrac{1}{a}(2a - |t|) & \text{for } a \leqslant |t| \leqslant 2a, \\ 0 & \text{for } |t| > 2a. \end{cases}$$

The function $\widetilde{V}_a(t)$ is now a Fourier transform of the Vallée–Poussin kernel

$$V_a(x) = \frac{1}{a} \frac{\cos(ax) - \cos(2ax)}{x^2}.$$

We have

$$\int_{\mathbb{R}} (1 - I(t)) \frac{f_{S_n}(t) - f_Y(t)}{t} \widetilde{h}(t) \widetilde{V}_a(t)\, \mathrm{e}^{-itx}\, \mathrm{d}t$$
$$= \big((F_{S_n} * h(x) - F_{S_n} * h * U_\Delta(x)) - (F_Y * h(x) - F_Y * h * U_\Delta(x))\big) * V_a(x),$$

where $\widetilde{h}(t)$ is a ch.f. of the c.d.f. h, and

$$U_\Delta(x) = \frac{1}{2\pi} \frac{\sin(\Delta x)}{x}.$$

(Note that the Fourier transform of U_Δ is the indicator function I.) Now, we obtain

$$\sup_x \left| \big((F_{S_n}(x) - F_{S_n} * U_\Delta(x)) * h(x) - (F_Y(x) - F_Y * U_\Delta(x)\big) * h * V_a(x) \right|$$
$$\leqslant \frac{d_{\delta,\gamma}(X, Y)}{n^{\frac{\gamma}{\alpha}-1}} \frac{(2a)^\gamma}{\gamma} \sqrt{2\pi}.$$

It is known (see, for example, [18]) that

$$\left| F_{S_n} * h(x) - F_{S_n} * h * V_a(x) \right| \leqslant \mathcal{E}_{F_{S_n} * h(x)}(a) \leqslant \mathcal{E}_h(a),$$

where $\mathcal{E}_F(a)$ is the order of the best approximation to the function F by entire functions of finite exponential type not greater than a. In our case, h has a bounded density function, so $\mathcal{E}_h(a) \leqslant c(h)/a$. Similarly, $|F_Y * h(x) - F_Y * h * V_a(x)| \leqslant c(h)/a$.

Let us recall a relation between norms of entire functions of finite exponential type (see [18], p. 125).

Suppose that $1 \leqslant p \leqslant p' \leqslant \infty$, *and let* $g \in L_p(\mathbb{R}^1)$ *be an entire function of exponential type* ν. *Then*

$$\|g\|_{L_{p'}(\mathbb{R}^1)} \leqslant 2\nu^{\frac{1}{p} - \frac{1}{p'}} \|g\|_{L_p(\mathbb{R}^1)}.$$

From this statement it follows that

$$\sup_x \left| (F_{S_n}(x) - F_Y(x)) * h * V_a * U_\Delta(x) \right| \leqslant 2\Delta a.$$

Combining our estimates we have

$$k_h(F_{S_n}, F_Y) \leqslant \inf_{a>0} \left(\sqrt{2\pi}\, \frac{d_{\delta,\gamma}(X, Y)(2a)^\gamma}{n^{\frac{\gamma}{\alpha}-1}\gamma} + 2\frac{c(h)}{a} + 2\Delta a \right)$$

for all $n \leqslant (\Delta/\delta)^\alpha$. □

Thus, the c.d.f. of the normalized sums of i.i.d. r.v.'s is close to the corresponding α-stable distribution for 'mid-size values' of n.

Theorem 2.1 shows that for ‘mid-size values’ of n the closeness of S_n to a strictly α-stable r.v. depends on the ‘middle part’ (‘body’) of the distribution of X. It gives (in some sense) a well-posed version of Central Limit Thorem.

Remark 2.2. Consider the example of radioactive decay and apply Theorem 2.1 to the centralized time moments (denote them by X_j). If Y is Gaussian, $\gamma = 3$, $\alpha = 2$, $\Delta = 10^{-10}$, $\delta = 10^{-30}$ we have that for $n \leqslant 10^{40}$ the following inequality holds:

$$k_h(F_{S_n}, F_Y) \leqslant \inf_{a>0} \left(\sqrt{2\pi}\, \frac{d_{10^{-30},3}(X, Y)(2a)^3}{3\sqrt{n}} + 2\frac{c(h)}{a} + 2 \times 10^{-10} a \right).$$

Here $d_{10^{-30},3}(X, Y) \leqslant 1$ (in view of the fact $|f_X(t) - f_Y(t)| \sim \frac{A^2}{(E_o^2+\Gamma^2)^2} t$, as $t \to 0$). So we have obtained a good normal approximation of $F_{S_n}(x)$ for ‘not too large’ values of n, namely, for $n \leqslant 10^{40}$. (If $c(h) \leqslant 1$ and n is of order 10^{40} then $k_h(F_{S_n}, F_Y)$ is of order 10^{-5}).

It is possible to obtain an analog of Central Pre-limit Theorem for Lévy distance

$$\begin{aligned} L(X, Y) = L(F_X, F_Y) \ = \ \inf\{\varepsilon\colon\ & F_X(x) \leqslant F_Y(x + \varepsilon) + \varepsilon, \\ & F_Y(x) \leqslant F_X(x + \varepsilon) + \varepsilon;\, x \in \mathbb{R}\} \end{aligned}$$

instead of Kolmogorov metric k_h.

THEOREM 2.2. *In conditions of Theorem* 2.1

$$L(S_n, Y) \leqslant \frac{\sqrt{2\pi}\, 2^\gamma}{\gamma} \frac{d_{\delta,\gamma}(X, Y)}{\Delta^{2\gamma/3}} \frac{1}{n^{\gamma/\alpha - 1}} + 6\Delta^{1/3}, \tag{2.1}$$

for all $n \leqslant (\Delta/\delta)^\alpha$.

Proof. For any positive η we have

$$L(S_n, Y) \leqslant k_h(S_n, Y) + \xi(\eta),$$

where $\xi(\eta) = \max\{2\eta, 1 - h(\eta), h(-\eta)\}$ (see [24], Lemma 1.5.2, p. 108; but we use other notations). For

$$h(x) = h_\eta(x) = \begin{cases} 0, & x \leqslant -\eta; \\ \dfrac{1}{2} + \dfrac{x}{\eta} + \dfrac{x^2}{2\eta^2}, & -\eta < x \leqslant 0; \\ \dfrac{1}{2} + \dfrac{x}{\eta} - \dfrac{x^2}{2\eta^2}, & 0 < x \leqslant \eta; \\ 1, & x > \eta \end{cases}$$

we obtain $\xi(\eta) = 2\eta$, and

$$L(S_n, Y) \leqslant \frac{\sqrt{2\pi}\, 2^\gamma}{\gamma} \frac{d_{\delta,\gamma}}{n^{\gamma/\alpha - 1}} a^\gamma + \frac{2}{\eta a} + 2\Delta a + 2\eta.$$

Choosing here $a = \Delta^{-2/3}$, $\eta = \Delta^{1/3}$ we finish the proof. □

Using the relations between uniform distance

$$\rho(X, Y) = \sup_{x\in\mathbb{R}} |F_X(x) - F_Y(x)|$$

and Lévy distance L (see [24, p. 107]) we obtain under conditions of the Theorem 2.1 that

$$\rho(S_n, Y) \leqslant \left(\frac{\sqrt{2\pi}2^\gamma}{\gamma}\frac{d_{\delta,\gamma}(X, Y)}{\Delta^{2\gamma/3}}\frac{1}{n^{\gamma/\alpha-1}} + 6\Delta^{1/3}\right)\left(1 + \sup_{x\in\mathbb{R}} p_Y(x)\right), \quad (2.2)$$

for all $n \leqslant (\Delta/\delta)^\alpha$, where $p_Y(x)$ is a density function of α-stable random variable Y.

3. Sums of a Random Number of Random Variables

Limit theorems for random sums of random variables have been studied by many specialists in probability, queueing theory, survival analysis, finance, econometric theory, etc.; we refer to [4, 9, 10, 12, 16, 17] and references therein.

We briefly recall the standard model: suppose X, X_j, $j \geqslant 1$, are i.i.d. r.v.'s and let $\{\nu_p, p \in \Delta \subset (0, 1)\}$ be a family of positive integer-valued random variables independent of the sequence of X's. Suppose that $\{\nu_p\}$ is such that there exists a *ν-strictly stable r.v. Y*, that is

$$Y \stackrel{d}{=} p^{1/\alpha}\sum_{j=1}^{\nu_p} Y_j,$$

where Y, Y_j, $j \geqslant 1$, are i.i.d. r.v.'s independent of ν_p, and $E\nu_p = 1/p$.

In Bunge [2], and Klebanov and Rachev [9] the authors independently obtained general conditions guaranteeing the existence of analogues of strictly stable distributions for sums of a random number of i.i.d. r.v.'s. For this type of a random summation model we can derive an analogue of Theorem 2.1.

THEOREM 3.1. *Let X, X_j, $j \geqslant 1$, be i.i.d. r.v.'s. Let $\tilde{S}_p = p^{1/\alpha}\sum_{j=1}^{\nu_p} X_j$. Suppose that $\tilde{Y}$ is a strictly ν-stable r.v. Let $\gamma > \alpha$, and $\Delta > \delta$ be arbitrary given positive constants, and let $p \geqslant (\delta/\Delta)^\alpha$ be an arbitrary positive number from* $(0, 1)$. *Then the following inequality holds*:

$$k_h\big(F_{\tilde{S}_p}, F_{\tilde{Y}}\big) \leqslant \inf_{a>0}\left(p^{\gamma/\alpha-1}\sqrt{2\pi}\frac{d_{\delta,\gamma}(X, \tilde{Y})(2a)^\gamma}{\gamma} + 2\frac{c(h)}{a} + 2\Delta a\right).$$

Proof. The proof is similar to that of Theorem 2.1. One only needs to use the following inequality

$$\begin{aligned} d_{c,\gamma}\big(\tilde{S}_p, \tilde{Y}\big) &\leqslant \sup_{|t|\geqslant c}\frac{\sum_{j=1}^{\nu_p}|f_X^n(p^{1/\alpha}t) - f_{\tilde{Y}}^n(p^{1/\alpha}t)|P(\nu_p = n)}{|t|^\gamma} \\ &\leqslant \sup_{|t|\geqslant c}\frac{|f_X(p^{1/\alpha}t) - f_{\tilde{Y}}(p^{1/\alpha}t)|E\nu_p}{|t|^\gamma} \\ &= p^{\gamma/\alpha-1}d_{cp^{1/\alpha},\gamma}(X, \tilde{Y}), \end{aligned}$$

at the beginning of the proof and then follow the arguments in the proof of Theorem 2.1. □

4. Local Pre-Limit Theorems and Their Applications to Finance*

Now we formulate our 'pre-limit' analogue of the classical local limit theorem.

THEOREM 4.1 (Local Pre-Limit Theorem). *Let X, X_j, $j \geqslant 1$, be i.i.d. r.v.'s having a bounded density function with respect to the Lebesgue measure, and $S_n = n^{-1/\alpha}\sum_{j=1}^{n} X_j$. Suppose that Y is a strictly α-stable random variable. Let $\gamma > \alpha$, $\Delta > \delta > 0$ and $n(\Delta/\delta)^\alpha$ be a positive integer not greater than $(\frac{\Delta}{\delta})^\alpha$. Then*

$$k_h(p_{S_n}, p_Y) \leqslant \inf_{a>0}\left(\sqrt{2\pi}\,\frac{d_{\delta,\gamma}(X,Y)(2a)^{\gamma+1}}{n^{\gamma/\alpha-1}(\gamma+1)} + 2\frac{c(h)}{a} + 2c(h)\Delta a\right),$$

where p_{S_n} and p_Y are the density functions of S_n and Y, respectively.

Thus the density function of the normalized sums of i.i.d. r.v.'s is close in smoothed Kolmogorov distance to the corresponding density of an α-stable distribution for 'mid-size values' of n.

The corresponding local pre-limit result for the sums of random number of random variables has the following form.

THEOREM 4.2 (Local Pre-Limit Theorem for Random Sums). *Let X, X_j, $j \geqslant 1$, be i.i.d. r.v.'s having bounded density function with respect to the Lebesgue measure. Let $\tilde{S}_\tau = \tau^{1/\alpha}\sum_{j=1}^{\nu_\tau} X_j$. Suppose that $\tilde{Y}$ is a strictly ν-stable random variable. Let $\gamma > \alpha$, and $\Delta > \delta > 0$, and $\tau \in [(\Delta/\delta)^\alpha, 1)$. Then the following inequality holds:*

$$k_h(p_{\tilde{S}_\tau}, p_{\tilde{Y}}) \leqslant \inf_{a>0}\left(\tau^{\gamma/\alpha-1}\sqrt{2\pi}\,\frac{d_{\delta,\gamma}(X,\tilde{Y})(2a)^{\gamma}}{\gamma} + 2\frac{c(h)}{a} + 2\Delta a\right).$$

Remark 4.1. Consider now our first example concerning Pareto-stable laws. Following the Mandelbrot [15] model for asset returns we view a daily asset return as a sum of a random number of tick-by-tick returns observed during the day. Following [9, 16, 17] we can assume that the total number of tick-by-tick returns during the day has a geometric distribution with a large expected value. In fact, the limiting distribution for geometric sums of random variables (when the expected value of the total number tends to infinity) is geo-stable [10]. Then, according to

* Note that in financial studies the fit of a theoretical distribution to the empirical one is often done in terms of the densities, rather than in terms of the corresponding c.d.f.'s. That is why, in our view, the local prelimit and limit theorems are of greater importance in comparison to the classical limit theorems when applied to financial studies.

our Theorem 4.2 the density function of daily returns is approximately geo-stable (in fact, it is ν-stable with a geometrically distributed ν).

5. Pre-Limit Theorem for Extremums

Let $X_1, \dots, X_n, \dots$ be a sequence of nonnegative i.i.d. random variables having the c.d.f. $F(x)$.

Denote

$$X_{1;n} = \min(X_1, \dots, X_n).$$

It is well-known that if $F(x) \sim ax^{\alpha}$ as $x \to 0$ then $F_n(x)$ (c.d.f. of $n^{1/\alpha} X_{1;n}$) tends to the cdf $G(x)$ of Weibull law, where

$$G(x) = \begin{cases} 1 - \mathrm{e}^{-ax^{\alpha}}, & \text{for } x > 0; \\ 0, & \text{for } x \leqslant 0. \end{cases}$$

The situation here is almost the same as in limit theorem for sums of random variables. It is obvious that the index α cannot be defined using empirical data on c.d.f. $F(x)$, and therefore *the problem of finding the limit distribution G is ill-posed*. Here we propose pre-limit version of corresponding limit theorem.

As an analogue of $d_{c,\gamma}$ we introduce another semi-distance between random variables X, Y:

$$\kappa_{c,\gamma}(X, Y) = \sup_{x>c} \frac{|F_X(x) - F_Y(x)|}{x^{\gamma}},$$

where F_X and F_Y are cdf's of nonnegative random variables X and Y.

THEOREM 5.1. *Let X_j, $j \geqslant 1$, be nonnegative i.i.d. r.v.'s and $X_{1;n} = \min(X_1, \dots, X_n)$. Suppose that Y is a random variable having the Weibull distribution*

$$G(x) = \begin{cases} 1 - \mathrm{e}^{-ax^{\alpha}}, & \textit{for } x > 0; \\ 0, & \textit{for } x \leqslant 0. \end{cases}$$

Let $\gamma > \alpha$ and $\Delta > \delta$ are arbitrary given positive constants, and $n < (\Delta/\delta)^{\alpha}$ be and arbitrary positive integer. Then

$$\sup_{x>0} \left|F_n(x) - G(x)\right| \leqslant \inf_{A>\Delta} \left(2\mathrm{e}^{-aA^{\alpha}} + 2\left(1 - \mathrm{e}^{-a\Delta^{\alpha}}\right) + \frac{A^{\gamma}}{n^{\gamma/\alpha - 1}} \kappa_{\delta,\gamma}(F, G) \right).$$

A little ough estimator under the conditions of the theorem and $\Delta < 1$ has the form

$$\sup_{x>0} \left|F_n(x) - G(x)\right| \leqslant \left(2 + \frac{1}{a^{\gamma/\alpha}} \left(\log \frac{1}{\varepsilon_n} \right)^{\gamma/\alpha} \right) \varepsilon_n + 2\left(1 - \mathrm{e}^{-a\Delta^{\alpha}}\right),$$

where

$$\varepsilon_n = \frac{1}{n^{\gamma/\alpha-1}}\kappa_{\delta,\gamma}(F,G).$$

To get this inequality it is sufficient to calculate instead the minimum the corresponding value for $A = (\frac{1}{a}\log\frac{1}{\varepsilon_n})^{1/\alpha}$.

Proof of Theorem 5.1. We have

$$\begin{aligned}\kappa_{\Delta,\gamma}(F_n,G) &= \kappa_{\Delta,\gamma}(F_n,G_n) = \sup_{x>\Delta}\frac{|F^n(x/n^{1/\alpha}) - G^n(x/n^{1/\alpha})|}{x^\gamma}\\ &\leqslant n\sup_{x>\Delta}\frac{|F(x/n^{1/\alpha}) - G(x/n^{1/\alpha})|}{x^\gamma} = \frac{1}{n^{\gamma/\alpha-1}}\kappa_{\Delta/n^{\gamma/\alpha-1},\gamma}(F,G)\\ &\leqslant \frac{1}{n^{\gamma/\alpha-1}}\kappa_{\delta,\gamma}(F,G)\end{aligned}$$

for $n \leqslant (\Delta/\delta)^\alpha$. So that

$$\kappa_{\Delta,\gamma}(F_n,G) \leqslant \frac{1}{n^{\gamma/\alpha-1}}\kappa_{\delta,\gamma}(F,G). \tag{5.1}$$

The inequality (5.1) shows that

$$\left|F_n(x) - G(x)\right| \leqslant x^\gamma\frac{1}{n^{\gamma/\alpha-1}}\kappa_{\delta,\gamma}(F,G) \tag{5.2}$$

holds for all $x \geqslant \Delta$. In particular

$$F_n(\Delta) \leqslant G(\Delta) + \Delta^\gamma\varepsilon_n.$$

Since $F_n(x) \leqslant F_n(\Delta)$ for $0 \leqslant x \leqslant \Delta$ then

$$\left|F_n(x) - G(x)\right| \leqslant 2G(\Delta) + \Delta^\gamma\varepsilon_n = 2\left(1 - \mathrm{e}^{-a\Delta^\gamma}\right) + \Delta^\gamma\varepsilon_n$$

for $0 \leqslant x \leqslant \Delta$.

For arbitrary $A > \Delta$ we have from (5.2)

$$\bar{F}_n(A) \leqslant \bar{G}(A) + A^\gamma\varepsilon_n$$

(where we use the notation $\bar{F}(x) = 1 - F(x)$) and therefore

$$\left|F_n(x) - G(x)\right| \leqslant 2\bar{G}(A) + A^\gamma\varepsilon_n$$

for $x \geqslant A$.

But from (5.2) we have

$$\sup_{\Delta<x<A}\left|F_n(x) - G(x)\right| \leqslant A^\gamma\varepsilon_n.$$

Combining the estimators for all values of x we finally get

$$\sup_{x>0} \left| F_n(x) - G(x) \right| \leqslant \inf_{A>\Delta} \left(2G(\Delta) + 2\bar{G}(A) + A^{\gamma} \varepsilon_n \right)$$

which complete the proof. □

6. Relations with Robustness of Statistical Estimators

Let $X, X_1, \ldots, X_n$ be a random sample from a population having c.d.f. $F(x, \theta)$, $\theta \in \Theta$ (which we shall call 'the model' here). For simplicity we shall further assume that $F(x, \theta)$ is c.d.f. of Gaussian law with θ mean and unit variance, so that $F(x, \theta) = \Phi(x - \theta)$ where $\Phi(x)$ is c.d.f. of standard normal law. Basing on the observations we have to construct an estimator $\theta^* = \theta^*(X_1, \ldots, X_n)$ of the θ-parameter.

The main point in the theory of robust estimation is that any proposed estimator should be insensitive (or weakly sensitive) to slight changes of underlying model, that is it should be *robust* [6].

For mathematical formalization of this we have to clarify two conceptions. The first one is the idea of how to express the notation of 'slight changes of underlying model' in quantitative form. And the second is the idea of the measurement of the quality of an estimator.

The most popular definition of the changes of the model in the theory of robust estimation is the following contamination scheme. Instead of the normal c.d.f. $\Phi(x)$ is considered $G(x) = (1 - \varepsilon)\Phi(x) + \varepsilon H(x)$, where $H(x)$ is an arbitrary symmetric c.d.f. Of course, for small values of $\varepsilon > 0$ the family $G(x - \theta)$ is close to the family $\Phi(x - \theta)$.

Sometimes the closeness of the families of c.d.f.'s is considered in terms of uniform distance between corresponding c.d.f.'s, or in terms of Lévy distance.

As to the measurement of the quality of an estimator then it is an asymptotic variance of the estimator.

Well known fact is that the minimum variance estimator for the parameter θ in 'pure' model $\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j$ is *non-robust*.

From our point of view, it is mostly connected *not* with the presence of contamination, but with the use of asymptotic variance as a loss function. Really, for not too large n we can apply our Theorem 2.1. It is easy to see that

$$d_{c,\gamma}\left(\Phi(x - \theta), G(x - \theta)\right) \leqslant 2\frac{\varepsilon}{c^{\gamma}}.$$

Suppose that $z_1, \ldots, z_n$ is a sample from the population with c.d.f. $G(x - \theta)$, and let $u_j = (z_j - \theta)$, $j = 1, \ldots, n$. Denote

$$S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} u_j = \sqrt{n}(\bar{z} - \theta).$$

For any $h(x)$ with a continuous density function, $\sup_x |h^{'}(x)| \leqslant 1$ we have

$$k_h\big(F_{S_n}, \Phi\big) \leqslant 2 \inf_{a>0} \left(\sqrt{2\pi}\,\frac{\varepsilon}{\delta^\gamma}\,\frac{(2a)^\gamma}{n^{\gamma/2-1}\gamma} + \frac{1}{a} + \Delta\cdot a\right).$$

Here $\gamma > 2$, $n \leqslant (\frac{\Delta}{\delta})^2$, and $\Delta > \delta > 0$ are arbitrary. It is not easy to find the inf over all positive values of a. Therefore we set $a = \Delta^{-1/2}$ to minimize the sum of two latest terms. Also we propose to find $\Delta = \varepsilon^c$ and $\delta = \varepsilon^{c_1}$ to have $\Delta^{1/2}\delta = \varepsilon^{1/\gamma}$. And, finally, we choose γ to maximize the degree c. Corresponding value is

$$\gamma = 2 + \sqrt{2/3},$$

and therefore

$$k_h\big(F_{S_n}, \Phi\big) \leqslant 2\left(\frac{\sqrt{2\pi}2^\gamma}{\gamma}\frac{1}{n^{1/\sqrt{6}}} + 2\varepsilon^{\frac{\sqrt{6}}{12+\sqrt{6}}}\right), \tag{6.1}$$

for all

$$n \leqslant \varepsilon^{-\frac{6}{12+7\sqrt{6}}}.$$

Here

$$\begin{aligned} \frac{\sqrt{2\pi}2^\gamma}{\gamma} &\cong 6.269467557, \\ \frac{1}{11} > \frac{\sqrt{6}}{12+\sqrt{6}} &\cong 0.08404082058 > \frac{1}{12}. \end{aligned}$$

From (6.1) we see, that (for very small ε) the properties of $\bar{z}$ as an estimator of θ do not depend on the tails of contaminating c.d.f. H for not too large values of the sample size. Therefore the traditional estimator for the location parameter of Gaussian law is robust for proper defined loss function. Let us note that the estimator of 'stability' does *not* depend on whether is c.d.f. $H(x)$ symmetric or not, though the assumption of symmetry is essential when the loss function coincides with asymptotic variance.

Of course, we can obtain corresponding estimator for both Lévy and uniform distances, but the order of 'stability' will be worse. For example, the Lévy distance estimator has the form

$$L\big(F_{S_n}, \Phi\big) \leqslant 2\left(\frac{\sqrt{2\pi}2^\gamma}{\gamma}\frac{1}{n^{\sqrt{3/10}}} + 3\varepsilon^{\frac{\sqrt{30}}{60+13\sqrt{30}}}\right)$$

for all

$$n \leqslant \varepsilon^{-\frac{10}{60+13\sqrt{30}}},$$

where

$$\gamma = 2 + \frac{\sqrt{30}}{5}.$$

We shall not give here the estimator for uniform distance.

One of possible objections is that the order of ‘stability’ in (6.1) is highly bad. But this circumstance is connected with ‘not proper’ choice of the distance between the distributions under consideration. It would be better to use $d_{c,\gamma}$ as a measure of closeness of corresponding model and real c.d.f.’s. Really, if

$$d_{\varepsilon,\gamma}\big(\Phi(x-\theta), G(x-\theta)\big) \leqslant \varepsilon,$$

and $c(h) \leqslant 1$ then

$$k_h\big(F_{S_n}, \Phi\big) \leqslant 4\left(\frac{2\sqrt{2\pi}}{n} + \varepsilon^{1/4}\right) \tag{6.2}$$

for all $n \leqslant \frac{1}{\varepsilon}$, which is better than (6.1).

7. Statistical Estimation for Nonsmooth Densities

Now we shall consider some relations between prelimit theorems for extremums and statistical estimation for nonsmooth densities. A typical example here gives a problem of estimation of the scale parameter for uniform distribution. Let us describe it in more details.

Suppose that $U_1, \ldots, U_n$ are i.i.d. random variables uniformly distributed over interval $(0, \theta)$. Basing on the data we have to estimate the parameter $\theta > 0$. It is known that the statistic

$$U_{n;n} = \max\{U_1, \ldots, U_n\}$$

is the best equivariant estimator for θ. Moreover, the distribution of $n(\theta - U_{n;n})$ tends to exponential law as n tends to infinity. In other words, the speed of convergence of $U_{n;n}$ to the parameter θ is $1/n$. But it is well-known that the speed of convergence of statistical estimator to ‘true’ value of the parameter is $1/\sqrt{n}$ in the case of smooth density function of the observations. More detailed formulations may be found in [7].

Our point here is that it is inpossible to verify basing on empirical observations does a density function have a discontinuity point or not. On the other hand, any c.d.f. having a density with point of discontinuity can be approximated (arbitrary closely) by c.d.f. having continuous density. But the speed of convergence for corresponding statistical estimators differs essentially ($1/n$ for the case of jump, and $1/\sqrt{n}$ in continuous case). It means that the problem of asymptotic estimation is *ill-posed*, and we have the situation very similar to that of summation of random variables.

Let now $X_1, \dots, X_n$ be a sample from population with c.d.f. $F(x/\theta)$, $\theta > 0$ ($F(+0) = 0$). Consider $X_{n;n}$ as an estimator for θ, and introduce

$$Z_j = \frac{\theta - X_j}{\theta}, \quad j = 1, \dots, n.$$

It is obvious that $Z_{1;n} = (\theta - X_{n;n})/\theta$. Therefore we can apply pre-limit theorem for minimums (see Theorem 5.1) to study the closeness of distribution of normalized estimator to the limit exponential distribution for pre-limit case. We have

$$\mathbb{P}_\theta\{Z_j < x\} = \mathbb{P}_\theta\{X_j > (1-x)\theta\} = 1 - F(1-x),$$

and we see that the c.d.f. of Z_j does not depend on θ. Let us denote by F_z the c.d.f. of Z_j. Denote by F_n c.d.f. of $nZ_{1;n}$, and by G – c.d.f. of exponential law $G(x) = 1 - \exp\{-x\}$ for $x > 0$. From Theorem 5.1 in the case of $\alpha = 1$ we obtain

$$\sup_x \left|F_n(x) - G(x)\right| \leqslant \inf_{A>\Delta} \left(2\mathrm{e}^{-A} + 2\big(1 - \mathrm{e}^{-\Delta}\big) + \frac{A^\gamma}{n^{\gamma-1}}\kappa_{\delta,\gamma}(F_z, G)\right) \quad (7.1)$$

for all $n \leqslant \frac{\Delta}{\delta}$.

Let us consider an examle, when c.d.f. of observations has the form $F(x) = x$ for $0 < x \leqslant a$, where a is a fixed positive number, and $F(x)$ is arbitrary for $x > a$. In this case it is easy to verify that

$$\kappa_{a,2} \leqslant \tfrac{1}{2}.$$

Choosing in (7.1) $\delta = a$, $\Delta = \frac{1}{4}\log\frac{1}{a}\sqrt{a}$, and $A = \frac{1}{2}\log\frac{1}{a}$ we obtain that

$$\sup_x \left|F_n(x) - G(x)\right| \leqslant \sqrt{a}\log\tfrac{1}{a}$$

for all $n < \frac{1}{4}\log(1/a)/\sqrt{a}$. In other words, the distribution of normalized estimator remains close to the exponential distribution for not too large values of the sample size, although F does not belong to the attraction domain of this distribution.

References

1. Arnold, B. C.: *Pareto Distributions*, I.C. Publ. House, Fairland, MD, 1983.
2. Bunge, J.: Composition semigroups and random stability, *Ann. Probab.* **24** (1996), 1476–1489.
3. DuMouchel, W.: Estimating the stable index α in order to measure tail thickness: A critique, *Ann. Statist*. **11** (1983), 1019–1031.
4. Gnedenko, B. V. and Korolev, V. Yu.: *Random Summation. Limit Theorems and Applications*, CRC Press, Boca Raton, 1996.
5. Hahn, M. G. and Klass, M. J.: Approximation of partial sums of arbitrary i.i.d. random variables and the precision of the usual exponential upper bound, *Ann. Probab.* **25** (1997), 1451–1470.
6. Huber, P. J.: *Robust Statistics*, Wiley, Chichester, 1981.
7. Ibragimov, I. A. and Khasminskii, R. Z.: *Asymptotic Theory of Estimation*, Nauka, Moskow, 1979 (in Russian).

8. Khalfin, L. A.: Contribution to the decay theory of a quasi-stationary state, *J. Experimenth. Theoret. Phys.* **6** (1958), 1053–1063.
9. Klebanov, L. B. and Rachev, S. T.: Sums of a random number of random variables an their approximations with ν – accompanying infinitely divisible laws, *Serdica* **22** (1996), 471–498.
10. Klebanov, L. B., Maniya, G. M. and Melamed, J. A.: A problem of Zolotarev and analogs of infinite divisible and stable distributions in a scheme for summing of a random number of random variables, *Theory Probab. Appl.* **29** (1984), 791–794.
11. Kolmogorov, A. N.: Some latest works on limit theorems in probability theory, *Vestnik MGU* **10** (1953), 28–39 (in Russian).
12. Kozubowski, T. J. and Rachev, S. T.: The theory of geometric stable distributions and its use in modeling financial data, *European J. Oper. Res.* **74** (1994), 310–324.
13. Lukacs, E.: *Characteristic Functions*, 2nd edn, Griffin, London, 1970.
14. Mandelbrot, B.: Variables et processus stochastiques de Pareto–Levy, et la repartition des revenus, *C.R. Acad. Sci. Paris* **23** (1959), 2153–2155.
15. Mandelbrot, B.: The Pareto–Levy law and the distribution of income, *Internat. Econ. Rev.* **1** (1960), 79–106.
16. Mittnik, S. and Rachev, S. T.: Modeling asset returns with alternative stable distributions, *Econ. Rev.* **12** (1993), 261–330.
17. Mittnik, S. and Rachev, S. T.: Reply to comments on "Modeling asset returns with alternative stable distributions", *Econometric Rev.* **12** (1993), 347–389.
18. Nikol'skii, S. M.: *Approximation of the Functions of Several Variables and Embedding Theorems*, Nauka, Moscow, 1977.
19. Pareto, V.: *Cours d'Economie Politique*, F. Rouge, Lausanne, Switzerland, 1897.
20. Petrosky, T. and Prigogine, I.: *Advances in Chemical Physics*, Vol. XCIX, Wiley, Chichester, 1997, pp. 1–120.
21. Rachev, S. T.: *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, 1991.
22. Wilkinson, S. R., Bharucha, C. F., Fisher, M. C., et al.: Experimental evidence for non-exponential decay in quantum tunnelling, *Nature* **387** (1997), 575–577.
23. Winter, R. G.: Evolution of a quasi-stationary state, *Phys. Rev.* **123** (1961), 1503–1507.
24. Zolotarev, V. M.: *Modern Theory of Summation of Independent Random Variables*, Nauka, Moscow, 1986.
25. Zolotarev, V. M.: *One-Dimensional Stable Distributions*, Nauka, Moscow, 1983 (in Russian). English translation: Amer. Math. Soc., Provindence, RI, 1986.