Pre-limit Theorems and Their Applications

Lev B Klebanov

Robust, 2014

Introduction and Motivating Examples

There exists a considerable debate about the applicability of limit theorems in probability theory because in practice one deals only with finite samples. Consequently, in the real-world, because one never deals with infinite samples, one can never know whether the underlying distribution is heavy tailed, or just has a long but truncated tail. Limit theorems are not robust with respect to truncation of the tail or with respect to any change from “light” to “heavy” tail, or vice versa. An approach to classical limit theorems that overcomes this problem is the “pre-limiting” approach. The advantage of this approach is that it does not rely on the tails of the distribution, but instead on the “central section” (or “body”) of a distribution. Therefore, instead of a limiting behavior when the number n of identical and independently distributed (i.i.d.) observations tends to infinity, a pre-limit theorem provides an approximation for distribution functions when n is “large” but not too “large.” The pre-limiting approach that we discuss in this chapter is more realistic for practical applications than classical central limit theorems.

Two Motivating examples

To motivate the use of the pre-limiting approach, we provide two examples.

Example 1: Pareto-Stable Laws More than 100 years ago Vilfredo Pareto 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 $C > 0$ and $\alpha > 0$. About 60 years later, Benoit Mandelbrot (1959, 1960) argued that stable laws should provide a more appropriate model for income distributions. After examining some income data, Mandelbrot made the following two claims:

1. 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 income follows a stable law (Lévy's stable law).
2. The tails of the Gaussian law are too thin to describe the distribution of income in typical situations.

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 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 evidence that shows that income distributions have Pareto tails with index α between 3 and 4, so the variance is finite. 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 of exponential distributions is the random time for radioactive decay. The exponential distribution is in the domain of attraction of the Gaussian law. It has been shown in quantum physics that the radioactive decay may not be exactly exponentially distributed (See Khalfin (1958), Wintner (1961), and Petrovsky and Prigogine (1997)). Experimental evidence supported that conclusion (see Wilkinson et al., (1997)) But then one faces the following paradox. Let $p(t)$ be the probability density that a physical system is in the initial state at moment $t \geq 0$. It is known that $p(t) = |f(t)|^2$, where

$$f(t) = \int_0^\infty \omega(E) \exp(iEt) dE,$$

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

$$\omega(E) = \frac{A}{(E - E_o)^2 + \Gamma^2}, \quad E \geq 0,$$

(see Zolotarev (1983a) and the references therein), where A is a normalizing constant, and E_o and Γ are the mode and the measure of dissipation of the system energy (with respect to E_o). For typical nonstable physical systems, the ratio Γ/E_o is very small (of order 10^{-15} or smaller). Therefore, the quantity

$$f(t) = e^{iE_o t}\frac{A}{\Gamma}\int_{-\frac{E_o}{\Gamma}}^{\infty}\frac{e^{i\Gamma ty}}{y^2+1}dy$$

differs from

$$f_1(t) = e^{iE_o t}\frac{A}{\Gamma}\int_{-\infty}^{\infty}\frac{e^{i\Gamma ty}}{y^2+1}dy = \pi e^{iE_o t}\frac{A}{\Gamma}e^{-t\Gamma}, \quad t>0,$$

by a very small value (of magnitude 10^{-15}). That is, $p(t) = |f(t)|^2$ is approximately equal to $\left(\frac{\pi A}{\Gamma}\right)^2 e^{-2t\Gamma}$, which gives approximately the classical exponential distribution as a model for decay.

On the other hand, it is equally easy to find the asymptotic representation of $f(t)$ as $t \to \infty$. Namely,

$$\int_{-\frac{E_o}{\Gamma}}^{\infty} \frac{e^{i\Gamma t y}}{y^2+1} dy =$$

$$\int_{-\arctan(\frac{E_o}{\Gamma})}^{\frac{\pi}{2}} e^{i\Gamma t \tan z} dz \sim -\frac{\cos^2(\arctan(\frac{E_o}{\Gamma}))}{it\Gamma} e^{-itE_o}.$$

Therefore,

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

where

$$A = \frac{1}{\int_0^{\infty} \frac{dE}{(E-E_o)^2+\Gamma^2}},$$

so that

$$p(t) \sim \frac{A^2}{(E_o^2+\Gamma^2)^2} \frac{1}{t^2}, \quad \text{as} \;\; t \to \infty. \tag{0.1}$$

Therefore, $p(t)$ belongs to the domain of attraction of a stable law with index $\alpha = 1$. Thus, if $T_j, j \geq 1$ are i.i.d. random variables 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 would expect under exponential decay), but diverges to infinity. Does this mean that the exponential approximation cannot be used anymore?

The two examples illustrate that the model based on the limiting distribution leads to an "ill-posed" problem in the sense that a small perturbation of the tail of the underlying distribution changes significantly the limit behavior of the normalized sum of random variables.

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

Principle idea

Here is the main idea. Suppose for simplicity that $X_1, X_2, \ldots, X_n$ are i.i.d. symmetric random variables whose distribution tail is heavy, but the "main body" looks to be similar to that of the Gaussian distribution. It seems natural to suppose that the behavior of the normalized sum

$$S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j$$

will be as following. For small values of n, it will be more or less arbitrary, and for growing values of n up to some number N, it becomes closer and closer to the Gaussian distribution (the tail does not play too essential a role). After the moment N, the distribution of S_n deviates from the Gaussian (the role of the tail is now essential).

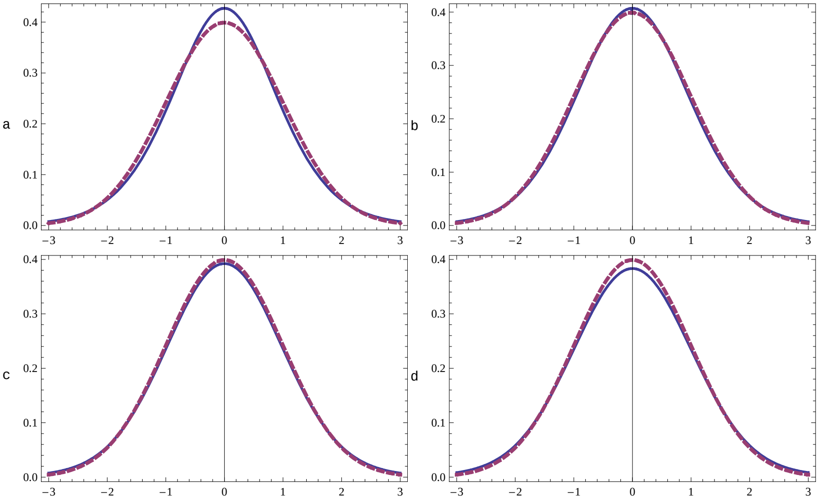
Let us illustrate this graphically. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with density function

$$p(x) = (1 - \varepsilon)q(x\sqrt{2}) + \varepsilon s(x).$$

Here $q(x) = \exp(-|x|)/2$ and $s(x) = 1/(\pi(1 + x^2))$ are the Laplacian and the Cauchy densities, respectively. Choose $\varepsilon = 0.01$. In panels *a* through *e* of Figure 1.1 we show the plot of the density of the sum

$$S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j$$

(the solid line) versus one of the density of the standard Gaussian distribution (the dashed line).


a
b
c
d
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3

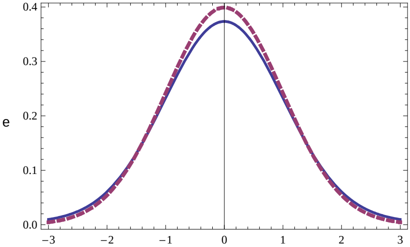


Figure: Density of a sum with different n versus Gaussian density

For $n = 5$ (panel a), we see that the densities are not too close to each other. When $n = 10$ (panel b), the two densities become closer to each other compared to when $n = 5$. They are almost identical when $n = 25$ (panel c). However, the two densities are not as close when $n = 50$ (panel d) and when $n = 100$ (panel e). Thus we see that the optimal N is about 25.

A very similar result is realized when the comparison is to a stable distribution. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with density function

$$p(x) = (1 - \varepsilon)q(2x) + \varepsilon s(x).$$

Here $q(x)$ is a density with ch.f. $(1 + |t|)^{-2}$, which belongs to a region of attraction of the Cauchy distribution and $s(x)$ is the density of the standard Gaussian distribution. We choose $\varepsilon = 0.03$.

In panels *a* and *b* of Figure 1.2 we show the plot of the density of the normalized sum

$$S_n = \frac{1}{n} \sum_{j=1}^{n} X_j$$

(the dashed line) versus one of the density of the Cauchy distribution (the solid line).
Panel *a* in the figure shows the two densities when $n = 5$. As can be seen, the densities are not too close to each other. However, as can be seen in panel b, the two densities become much closer to each other when $n = 50$

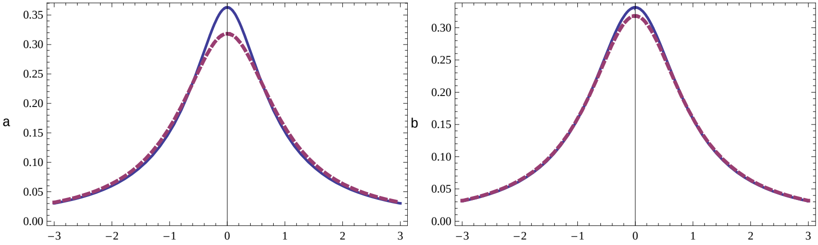


Figure: Density of a sum for various n (solid line) versus *Cauchy* density (dashed line)

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

$$d_{c,\gamma}(X,Y) = \sup_{|t|\geq c} \frac{|f_X(t) - f_Y(t)|}{|t|^\gamma}. \tag{0.2}$$

Here and in what follows F_X and f_X stand for the cumulative distribution function (c.d.f.) and the characteristic function (ch.f.) 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 random variables for which $d_{0,\gamma}(X, Y)$ is finite. Next, recall that Y is a strictly α-stable random variable. If for every positive integer n

$$Y_1 \stackrel{d}{=} U_n := \frac{Y_1 + \cdots + Y_n}{n^{1/\alpha}}, \tag{0.3}$$

where $\stackrel{d}{=}$ stands for equality in distribution and the Y_j's, $j \geq 1$, are i.i.d. copies of Y.

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

$$d_{0,\gamma}(S_n, Y) = d_{0,\gamma}(S_n, U_n)$$

$$= \sup_t \frac{|f_X^n(t/n^{1/\alpha}) - f_Y^n(t/n^{1/\alpha})|}{|t|^\gamma}$$

$$\leq n \sup_t \frac{|f_X(t/n^{1/\alpha}) - f_Y(t/n^{1/\alpha})|}{|t|^\gamma} = \frac{1}{n^{\gamma/\alpha - 1}} d_{0,\gamma}(X, Y).$$

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 about 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 can 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 providing answers to the problems addressed in Examples 1 and 2 in Section 1

Central Pre-Limit Theorem

In our Central Pre-Limit Theorem we shall analyze the closeness of the sum S_n to a strictly α-stable random variable Y in terms of the following Kolmogorov metric, defined for any c.d.f.'s F and G as follows:

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

Here, $*$ stands for convolution, and the "smoothing" function $h(x)$ is a fixed c.d.f. with a bounded continuous density function, $\sup_x |h^{'}(x)| \leq c(h) < \infty$. The metric k_h metrizes the weak convergence in the space of c.d.f.'s. We have the following pre-limit theorem.

Theorem (Central Pre-Limit Theorem)

Let $X, X_j, j \geq 1$, be i.i.d. random variables and $S_n = n^{-1/\alpha} \sum_{j=1}^{n} X_j$. Suppose that Y is a strictly α-stable random variable. Let $\gamma > \alpha$ and $\Delta > \delta$ be arbitrary given positive constants and let $n \leq (\frac{\Delta}{\delta})^{\alpha}$ be an arbitrary positive integer. Then

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

If $\Delta \to 0$ and $\Delta/\delta \to \infty$, then n can be chosen large enough so that the right-hand-side of the above bound is sufficiently small, and 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.

The c.d.f. of a normalized sum of i.i.d. random variables is close to the corresponding α-stable c.d.f. for "mid-size values" of n. We also see that for these values of n, the closeness of S_n to a strictly α-stable random variable depends on the "middle part" ("body") of the distribution of X.

Remark *Consider our example of radioactive decay and apply Theorem 0.1 to the centralized time moments, denoted by* X_j. *If* Y *is Gaussian,* $\gamma = 3$, $\alpha = 2$, $\Delta = 10^{-15}$, $\delta = 10^{-30}$, *then for* $n \le 10^{30}$ *the following inequality holds:*

$$k_h(F_{S_n}, F_Y) \le$$

$$\le \inf_{a>0} \left(\sqrt{2\pi} \frac{d_{10^{-30},3}(X, Y)(2a)^3}{3\sqrt{n}} + 2\frac{c(h)}{a} + 2 \cdot 10^{-10} a \right).$$

Here, $d_{10^{-30},3}(X, Y) \le 1$ *in view of the fact that*

$$|f_X(t) - f_Y(t)| \sim \frac{A^2}{(E_o^2 + \Gamma^2)^2} t, \quad \text{as } t \to 0.$$

Thus, we obtain a rather good normal approximation of $F_{S_n}(x)$ for "not too large" values of n ($n \leq 10^{40}$). If $c(h) \leq 1$ and n is of order 10^{40}, then $k_h(F_{S_n}, F_Y)$ is of order 10^{-5}.

Sums of a Random Number of Random Variables

Limit theorems for random sums of random variables have been studied by many specialists in such fields as probability theory, queueing theory, survival analysis, and financial econometric theory. We briefly recall the standard model: suppose $X, X_j, j \geq 1$, are i.i.d. random variables 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 random variable* Y, that is

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

where $Y, Y_j, j \geq 1$, are i.i.d. random variables independent of ν_p, and $E\nu_p = 1/p$.

Bunge (1996) and Klebanov and Rachev (1996) independently obtained general conditions guaranteeing the existence of analogues of strictly stable distributions for sums of a random number of i.i.d. random variables. For this type of a random summation model, we can derive an analogue of Theorem 0.1.

Theorem

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

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

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

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

Theorem (Local Pre-Limit Theorem)

Let $X, X_j, j \geq 1$, be i.i.d. random variables 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 $n(\frac{\Delta}{\delta})^{\alpha}$ be a positive integer not greater than $(\frac{\Delta}{\delta})^{\alpha}$. Then

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

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. random variables 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 (Local Pre-Limit Theorem for Random Sums)

Let $X, X_j, j \geq 1$, be i.i.d. random variables 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 $\tau \in [(\frac{\Delta}{\delta})^\alpha, 1)$. Then the following inequality holds:

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

Remark *Consider now our first example in Section 1 concerning Pareto-stable laws. Following the Mandelbrot (1960) 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 trading day. We can assume that the total number of tick-by-tick returns during the trading 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. Then, according to Theorem 1.4 from Klebanov, Rachev, Kozubowskii (2006), the density function of daily returns is approximately geo-stable (in fact, it is* ν*-stable with a geometrically distributed* ν*).*

Pre-Limit Theorem for Extremums

Let $X_1, \dots, X_n, \dots$ be a sequence of non-negative 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 c.d.f. $G(x)$ of the Weibull law, where

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

The situation here is almost the same as in the 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 the pre-limit version of the 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 c.d.f.'s of random variables $X, Y > 0$.

Theorem

Let X_j, $j \geq 1$, be non-negative i.i.d. random variables and $X_{1;n} = \min(X_1, \ldots, X_n)$. Suppose that Y is a random variable having the Weibull distribution

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

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

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

A little more rough estimator under the conditions of Theorem 1.5 and $\Delta < 1$ has the form

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

where

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

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

Relations with Robustness of Statistical Estimators

Let $X, X_1, \dots, X_n$ 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 a 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. One uses the observations $X_1, \dots, X_n$ to construct an estimator $\theta^* = \theta^*(X_1, \dots, X_n)$ 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 the underlying model; that is, it should be *robust*.

For mathematical formalization of this, we have to clarify two notions. 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. It is a well known fact that the minimum variance estimator for the parameter θ in a "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. For not too large n, we can apply Theorem 0.1.

It is easy to see that

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

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, $\sup_x |h^{'}(x)| \leq 1$, we have

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

Here $\gamma > 2$, $n \leq \left(\frac{\Delta}{\delta}\right)^2$, and $\Delta > \delta > 0$ are arbitrary. It is not easy to find the infimum over all positive values of a. Therefore, we set $a = \Delta^{-\frac{1}{2}}$ to minimize the sum of the two last 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.

The corresponding value is

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

and therefore

$$k_h(F_{S_n}, \Phi) \leq 2 \left(\frac{\sqrt{2\pi} 2^\gamma}{\gamma} \frac{1}{n^{1/\sqrt{6}}} + 2\varepsilon^{\frac{\sqrt{6}}{12+\sqrt{6}}} \right), \qquad (0.4)$$

for all

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

Here

$$\frac{\sqrt{2\pi} 2^\gamma}{\gamma} \cong 6.269467557,$$

$$\frac{1}{11} > \frac{\sqrt{6}}{12+\sqrt{6}} \cong 0.08404082058 > \frac{1}{12}.$$

Now 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 the Gaussian law is robust for a properly defined loss function. Note that the estimator of "stability" does not depend on whether c.d.f. $H(x)$ is symmetric or not, though the assumption of symmetry is essential when the loss function coincides with asymptotic variance.

Of course, we can obtain a 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

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

for all

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

where

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

We shall not provide here the estimator for uniform distance.

One possible objection is that the order of “stability” in our inequality is very bad. On the one hand, our estimators are not precise. On the other hand, it is related to the “improper” choice of the distance between the distributions under consideration. It would be better to use $d_{c,\gamma}$ as a measure of closeness of the corresponding model and real c.d.f.'s. If

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

and $c(h) \leq 1$, then

$$k_h(F_{S_n}, \Phi) \leq 4 \left(\frac{2\sqrt{2\pi}}{n} + \varepsilon^{\frac{1}{4}} \right) \tag{0.5}$$

for all $n \leq \frac{1}{\varepsilon}$, which is superior to previous inequality.

Probably, the estimator of stability is better for other type of distances. We can support this position with numerical examples. Namely, let $X_1, X_2, \dots, X_n$ be i.i.d. random variables distributed as a mixture of the standard Gaussian distribution (with weight $1-\varepsilon$) and Cauchy distribution (weight ε). The uniform distance between distribution $F(x, n, \varepsilon)$ of the normalized sum

$$S_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j$$

for $\varepsilon = 0.01$, $n = 50$ and the standard Gaussian distribution is approximately 0.014. For $\varepsilon = 0.02$, $n = 50$, this distance is about 0.027.

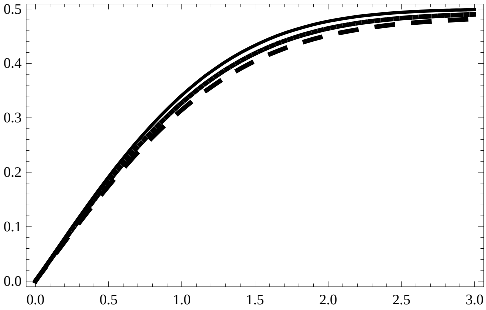


Figure: Plots of distributions of normalized sums

Figure 3 provides graphs of $F(x, n, \varepsilon) - 0.5$ for $n = 50$ and $\varepsilon = 0$ (solid line), $\varepsilon = 0.01$ (dashed line, short intervals), and $\varepsilon = 0.02$ (dashed line, long intervals).

We propose the use of models that are close to each other in terms of weak distances. Therefore, we cannot use such loss functions like the quadratic one because the risk of one estimator can become infinite. Therefore, we have to discuss possible choices for the losses. This is a major separate problem in statistics, and we refer the reader to Kakosyan, Klebanov, and Melamed (1984b).

Statistical Estimation for Non-Smooth Densities

Now we shall consider some relations between pre-limit theorems for extremums and statistical estimation for non-smooth densities. A typical example here is a problem of estimation of the scale parameter for a uniform distribution. Let us describe it in more detail.

Suppose that $U_1, \ldots, U_n$ are i.i.d. random variables uniformly distributed over interval $(0, \theta)$. Based 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 $\frac{1}{n}$. But it is known that the speed of convergence of a statistical estimator to the "true" value of the parameter is $\frac{1}{\sqrt{n}}$ in the case where there is a smooth density function of the observations. More detailed formulations may be found in Ibragimov and Khasminskii (1979).

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

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

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

It is obvious that $Z_{1;n} = \frac{\theta - X_{n;n}}{\theta}$. Therefore, we can apply the pre-limit theorem for minimums to study the closeness of the distribution of the normalized estimator to the limit exponential distribution for the 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 the exponential law $G(x) = 1 - \exp\{-x\}$ for $x > 0$. From Theorem for minimums in the case of $\alpha = 1$, we obtain

$$\sup_x |F_n(x) - G(x)| \le \inf_{A>\Delta} \left(2e^{-A} + 2(1 - e^{-\Delta}) + \frac{A^\gamma}{n^{\gamma-1}} \kappa_{\delta,\gamma}(F_z, G) \right) \tag{0.6}$$

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

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

$$\kappa_{a,2} \leq \frac{1}{2}.$$

Choosing in (0.6) $\delta = a$, $\Delta = \frac{1}{4}\log\frac{1}{a}\sqrt{a}$, and $A = \frac{1}{2}\log\frac{1}{a}$, we obtain that

$$\sup_x |F_n(x) - G(x)| \leq \sqrt{a}\log\frac{1}{a}$$

for all $n < \frac{1}{4}\frac{\log\frac{1}{a}}{\sqrt{a}}$. 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.

Let us now give some results of numerical simulations. We simulated $m = 50$ samples of the size $n = 1000$ from two populations. The first one is uniform on $(0, 1)$, and the second has the following distribution function

$$F(x,\varepsilon) = \begin{cases} 0, & \text{for} \quad x < 0, \\ x, & \text{for} \quad \leq x < 1-\varepsilon, \\ 1 - \frac{1}{\varepsilon^{1/4}}(1-x)^{5/4}, & \text{for} \quad 1-\varepsilon \leq x < 1, \\ 1, & \text{for} \quad x \geq 1 \end{cases}$$

with $\varepsilon = 0.005$. So, we had i.i.d. random variables $Y_{i,j}$, $i = 1, \dots, n$; $j = 1, \dots, m$ with uniform $(0, 1)$ distribution, and i.i.d. $X_{i,j}$ with distribution $F(x, \varepsilon)$. Denote $V_j = \max_i Y_{i,j}$ and $U_j = \max_i X_{i,j}$.

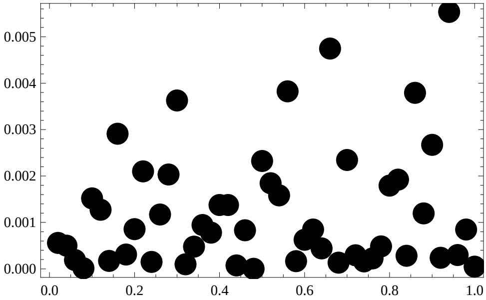


Figure: Simulated points $(j/m, 1 - V_j)$

In Figure 4 the simulated points $(j/m, 1 - V_j)$ are shown. The values $1 - V_j$ are identical to those of the difference between the true value of the scale parameter and the value of the estimator for the "true" model.

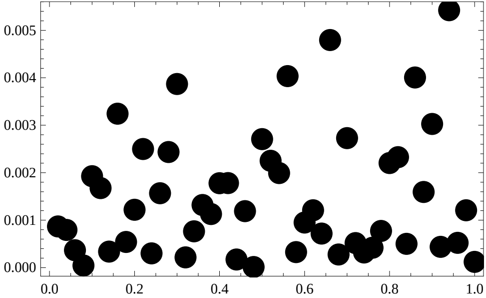


Figure: Simulated points $(j/m, 1 - U_j)$

In Figure 5 the simulated points $(j/m, 1 - U_j)$ are shown. The values $1 - U_j$ are identical to those of the difference between true value of the scale parameter and the value of the estimator for "perturbed" model. Comparing Figures 5 and 4 we can see that the simulated results are very similar.

We can also compare empirical distributions. We simulated $m = 5000$ samples of the size $n = 200$ each from the same populations as before. Now we consider normalized values of the differences between true value of the parameter and its statistical estimators: $n(1 - V_j)$ for the "pure" model, and $n(1 - U_j)$ for the "perturbed" model. Averaging over all $m = 5000$ samples, we find empirical distributions of the estimator in both models.

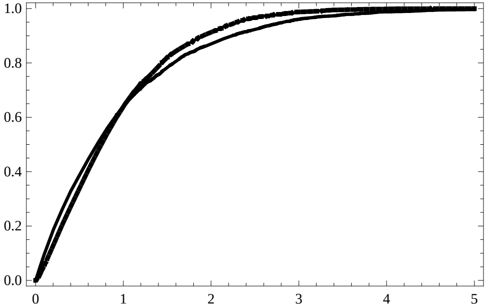


Figure: Graphs of distribution functions of the normalized estimators

Figure 6 shows the graphs of distributions of the normalized estimator for the “pure” (solid line) and for the “perturbed” models (dashed line). Of course, the agreement is rather good.

Thank yoou for your attention