# Ill-Posed Problems in Probability

# Preface

Here we consider so called *ill-posed* problems in statistics and probability theory. Ill-posed problems are usually understood as certain results where small changes in the assumptions lead to arbitrary large changes in the conclusions. Such results are not very useful for practical applications, where the presumptions usually hold only approximately (and even the slightest departure from the assumed model may produce an uncontrollable shift in the outcome). Presumably, the ill-posedness of certain practical problems is due to the lack of their precise mathematical formulation. Consequently, we can deal with such problems by replacing a given ill-posed problem with another, well-posed problem, which in some sense is "close" to the original one.

Our goal is to show that ill-posed problems are not a curiosity in the contemporary theory of mathematical statistics and probability. On the contrary, such problems are quite common, and the majority of classical results fall into this class. Our objective is to identify problems of this type, and re-formulate them more correctly. Thus, we shall propose alternative (more precise in the above sense) versions of numerous classical theorems in the theory of probability and mathematical statistics. In addition, we shall consider some non-standard problems from this point of view. Let us mention several examples of ill-posed problems.

First of all, the classical Central Limit Theorem, as well as the corresponding limit theorem for convergence to a stable law, are both examples of an ill-posed problem. Indeed, an arbitrary small perturbation (in the uniform metric) of the tail of the distribution leads to a shift of the domain of attraction: a normal domain of attraction may convert to a stable one, and vice versa. Corrected versions of these theorems were proposed independently by Nagaev and Klebanov. The main idea here is to replace the limiting distribution with an approximation of the *pre-limiting* distribution with a larger (but not too much larger) number of random variables in the sum.

The second example comes from extreme value theory and concerns limit theorems for extremal order statistics for i.i.d. random variables. The setting here is analogous to that for limit theorems for sums of i.i.d. random variables. More precisely, it is well-known that the limiting distribution of an appropriately normalized minimum of non-negative i.i.d. random variables is Weibull. The parameters of the limiting Weibull distribution depend on the rate of convergence to zero of the relevant distribution function. An arbitrary small changes (in the uniform metric) of the distribution function may affect this rate quite severely, and thus the problem of finding the exact limiting distribution appears to be ill-posed.

The third example of an ill-posed problem is the classical problem of estimating the location parameter of a normal distribution with known standard deviation. If the distribution of the measurement error is indeed a Gaussian one, then the optimal equivariant estimator of the location parameter is provided by the sample mean. However, if the sample may be contaminated with observations from a heavy-tail distribution and we are using the variance of the limiting distribution as the loss function, then the sample mean becomes unacceptable, as its variance may in general be infinite. This example has led to the theory of robust estimation, see, e.g., Huber and Hampel. It is clear that the above problem is closely connected with the correct formulation of the problem of finding the limiting distribution for sums of i.i.d. random variables. Following the recommendations of Klebanov, Rachev and Szekely, we shall consider approximations of the pre-limiting distribution. In this case we can not utilize the variance of the limiting distribution as the loss function. Thus, a corrected formulation of the problem of estimating the location parameter is two-folded, involving the pre-limiting approach as well as the issue of choosing an appropriate loss function.

The forth example is connected with statistical estimation of parameters when the underlying density has discontinuities (jumps). The typical characteristics of this example can be illustrated with the problem of estimating the scale parameter $\theta$ of the uniform distribution on the interval $(0, \theta)$. Here, if $X_1, \ldots, X_n$ is a random sample from this distribution, then the maximum, $X_{n:n} = \max(X_1, \ldots, X_n)$, is a consistent estimator for $\theta$. In fact the normalized sequence $n(X_{n:n} - \theta)$ has a non-singular limiting distribution as $n \to \infty$. However, when we replace the uniform distribution with another one, which is smooth and arbitrarily close to it (in the uniform metric on the space of distribution functions), then $X_{n:n}$ along with every other estimator will no longer be consistent. The corresponding rate of convergence will now be not larger than $1/\sqrt{n}$, that is, the normalizing constant $n$ needs to be replaced by $\sqrt{n}$. Thus, we again end up with an ill-posed problem. Its corrected version is based on the replacement of the limiting distribution of the normalized sequence by the pre-limiting distribution.

Finally, our last example is provided by the problem of specifying the distribution using a finite number of values of certain functionals. The examples of such functionals include the moments, or the Radon transformation of the original distribution. The latter is particularly common in the area of computer tomography. The proof of the ill-posedness here follows from an interesting example, discussed in Lagarrias, Kempermann, Shepp and Reeds. The corrected versions appeared in Khalfin and Klebanov. Some applications to quantum mechanics are discussed in Khalfin and Klebanov, Klebanov and Rachev.

Thus, we see the ill-posed problems appear quite frequently in statistics and probability theory.
The layout of the talk would correspond to the above examples in probability theory (not in Statistics).

# Limit Theorems and Ill-posed Problems

**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. 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. 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 $\alpha$ 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.[1] Recently, new experimental evidence supported that conclusion (see Wilkinson et al., (1997)).

[1] See Khalfin (1958), Wintner (1961), and Petrovsky and Prigogine (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[2] 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 $\Gamma$ are the mode and the measure of dissipation of the system energy (with respect to $E_o$). For typical nonstable physical systems, the ratio $\Gamma / E_o$ is very small (of order $10^{-15}$ or smaller).

[2] See, for example, Zolotarev (1983a, p. 42).

Therefore, the quantity

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

differs from

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

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

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

Therefore, $p(t)$ belongs to the domain of attraction of a stable law with index $\alpha = 1$. Thus, if $T_j, j \geq 1$ 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).

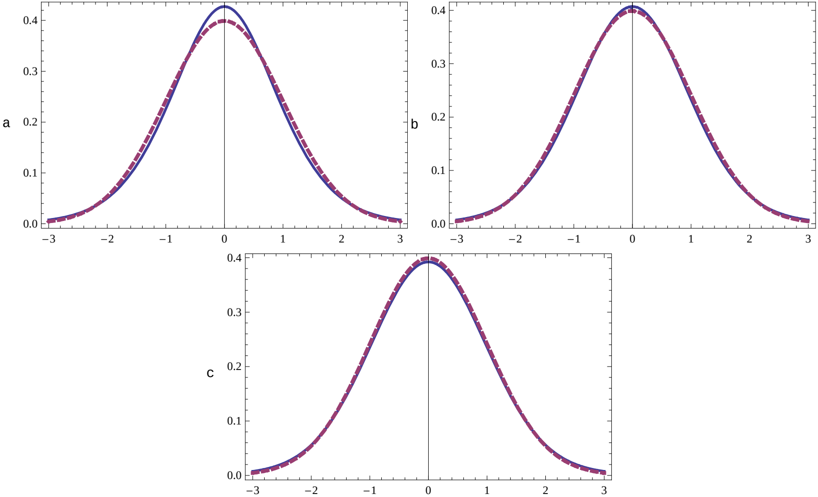
Let us illustrate this graphically. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with density function

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

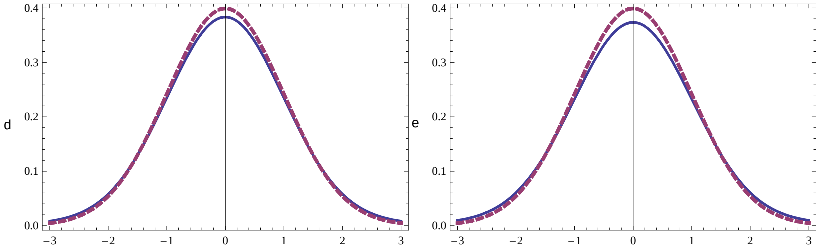
Here $q(x) = \exp(-|x|)/2$ and $s(x) = 1/(\pi(1 + x^2))$ are the Laplacian and the Cauchy densities, respectively. Choose $\varepsilon = 0.01$. In panels *a* through *e* of Figure 1.1 we show the plot of the density of the sum

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

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




Density of a sum with different $n$ versus Gaussian density ($n = 5$, $n = 10$, $n = 25$)




Density of a sum with different $n$ versus Gaussian density ($n = 50$, $n = 100$).

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

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

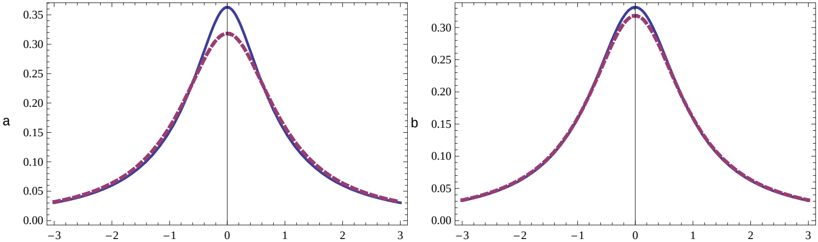
Here $q(x)$ is a density with ch.f. $(1 + |t|)^{-2}$, which belongs to a region of attraction of the Cauchy distribution and $s(x)$ is the density of the standard Gaussian distribution. We choose $\varepsilon = 0.03$.

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 $n = 50$




Density of a sum for various $n$ (solid line) versus *Cauchy* density (dashed line), $n = 5$, $n = 50$.

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

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

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[3].

[3]See Zolotarev (1986) and Rachev (1991).

Next, recall that $Y$ is a strictly $\alpha$-stable random variable. 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 the $Y_j$'s, $j \geq 1$, are i.i.d. copies of $Y$[4].

[4] See Zolotarev (1983a) and Lukacs (1969).

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[5]

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

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

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

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 $\alpha$-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 $d_{0,\gamma}(X, Y) = \infty$.

[5] See Zolotarev (1983a).

Our goal is to analyze the closeness of the sum $S_n$ to a strictly $\alpha$-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.

# Central Pre-Limit Theorem

In our Central Pre-Limit Theorem we shall analyze the closeness of the sum $S_n$ to a strictly $\alpha$-stable random variable $Y$ in terms of the following Kolmogorov metric,[6] 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. The following central pre-limit theorem appeared in Klebanov et al. (1999).

[6] See Kolmogorov (1953) and Rachev (1991).

### 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 $\alpha$-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*

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

### Remark

*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* $\alpha$*-stable law. This result, of course, includes the central limit theorem for weak distance.*

# Proof of Theorem.

For $\gamma > \alpha$,

$$d_{c,\gamma}(S_n, Y) = d_{c,\gamma}(S_n, T_n)$$

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

For any $\Delta > \delta$ and for all $n \leq (\frac{\Delta}{\delta})^\alpha$, we have then

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

The above relation can be rewritten in the form

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

Denote by $\mathbb{I}(t)$ the indicator function of the interval $[-\Delta, \Delta]$. Then,

$$\frac{1}{|t|} \left|(1 - \mathbb{I}(t)) f_{S_n}(t) - (1 - \mathbb{I}(t)) f_Y(t)\right| \leq \frac{|t|^{\gamma-1}}{n^{\frac{\gamma}{\alpha}-1}} d_{\delta,\gamma}(X, Y).$$

For any $a > 0$ define

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

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

$$\int_{\mathbb{R}} (1 - \mathbb{I}(t)) \frac{f_{S_n}(t) - f_Y(t)}{t} \widetilde{h}(t) \widetilde{V}_a(t) e^{-itx} dt$$

$$= ((F_{S_n} * h(x) - F_{S_n} * h * U_\Delta(x)) - (F_Y * h(x) - F_Y * h * U_\Delta(x))) * V_a(x$$

where $\widetilde{h}(t)$ is the ch.f. corresponding to the c.d.f. $h$ and

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

(Note that the Fourier transform of $U_\Delta$ is the indicator function $\mathbb{I}$.)

We now obtain

$$\sup_x |((F_{S_n}(x) - F_{S_n} * U_\Delta(x)) * h(x) - (F_Y(x) - F_Y * U_\Delta(x)) * h * V_a(x)|$$

$$\leq \frac{d_{\delta,\gamma}(X,Y)}{n^{\frac{\gamma}{\alpha}-1}} \frac{(2a)^\gamma}{\gamma} \sqrt{2\pi}.$$

It is known[7] that

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

where $\mathcal{E}_F(a)$ is the order of the best approximation of the function $F$ by entire functions of finite exponential type $a$.

[7] See Nikolskii (1977).

In our case, $h$ has a bounded density function, so $\mathcal{E}_h(a) \leq c(h)/a$. Similarly, $|F_Y * h(x) - F_Y * h * V_a(x)| \leq c(h)/a$.
From a well-known relation between norms of entire functions of finite exponential type (See, Nikolskii (1977, p. 131)), it follows that

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

Combining our estimates, we have

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

for all $n \leq (\frac{\Delta}{\delta})^\alpha$. $\square$
Thus, the c.d.f. of a normalized sum of i.i.d. random variables is close to the corresponding $\alpha$-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 $\alpha$-stable random variable depends on the "middle part" ("body") of the distribution of $X$.

## Remark

*Consider our example of radioactive decay and apply Theorem 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 $n \leq 10^{30}$ the following inequality holds:*

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

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

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

*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 $10^{-5}$.*

# 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 a c.d.f. of Gaussian law with $\theta$ 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 $X_1, \ldots, X_n$ to construct an estimator $\theta^* = \theta^*(X_1, \ldots, X_n)$ of the $\theta$-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*.[8]

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.

[8] See Huber (1981).

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 $\theta$ 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 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

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

Here $\gamma > 2$, $n \le \left(\frac{\Delta}{\delta}\right)^2$, 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 $\gamma$ to maximize the degree $c$. The corresponding value is

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

and therefore

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

for all

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

Here

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

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

We see that (for very small $\varepsilon$) the properties of $\bar{z}$ as an estimator of $\theta$ 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” 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

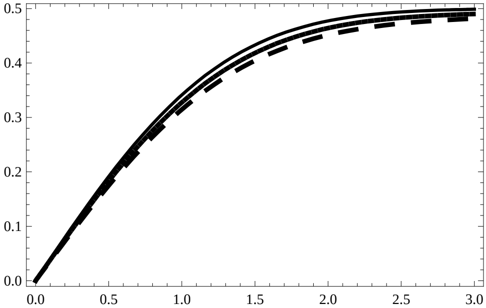
$$k_h(F_{S_n}, \Phi) \leq 4\left(\frac{2\sqrt{2\pi}}{n} + \varepsilon^{\frac{1}{4}}\right)$$

for all $n \leq \frac{1}{\varepsilon}$.

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 $\varepsilon$). 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.




Plots of distributions of normalized sums

This Figure 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 to Kakosyan, Klebanov and Melamed (1984b) and to Klebanov, Rachev and Fabozzi (2009).

# Ill-Posed Problems in Computer Tomography

Now we will consider certain problems of computer tomography. We will be concerned with studying the closeness of probability distributions with common marginal distributions, or with some common moments. Our methods are similar to those of the classical moment problem.

# The Radon Transform and its Applications to Computer Tomography

The Radon Transform and its Applications to Computer Tomography

The traditional methods of computer tomography are useful in many branches of science, medicine, and technology.

Suppose we have an object, possibly a human body, through which an $X$-ray passes, and we measure the intensities of the ray at both the input $I_i$ and the output $I_0$. It is known, that under some conditions imposed on the object and the intensity of the input ray,

$$\log(\frac{I_i}{I_0}) = \int_L p(x) dL,$$

where the integral of the density$p(x)$ at position $x$ is taken along $L$, the straight line that the $X$-ray follows through the body. We should like to reconstruct the density$p(x)$ of the body based on the line integrals calculated on some or all of the straight lines.

The transformation from the density $p(x)$ to the set of all its line integrals, considered as functions of the parameters of the line, is the *Radon transform* introduced in 1917 by Radon. Radon gave the formulae for the inversion of the transform and proved the uniqueness of the reconstruction of the density $p(x)$ from the transform.

More precisely, an *$n-$dimensional Radon transform* **R** maps a function given on $\mathbb{R}^n$ into the set of all its integrals over hyperplanes in $\mathbb{R}^n$:

$$\mathbf{R}f(\theta, s) = \int_{\langle x,\theta\rangle = s} f(x)dx = \int_{\theta^\perp} f(s\theta + y)dy.$$

Here the integral is taken over the hyperplane perpendicular to the vector $\theta$ and situated at the distance$s$ from the origin. Similarly, the *$X-$ray transform* **P** maps a function given on $\mathbb{R}^n$ into the set of its line integrals:

$$\mathbf{P}f(\theta, x) = \int_{-\infty}^{\infty} f(x + t\theta)dt.$$

This integral is taken over the straight line through the point $x$ in the direction $\theta$.

It is interesting to note that the Radon transform allows for a unique reconstruction of any probability measure on an $n$-dimensional Euclidean spacefrom the probabilities of half spaces. This is essentially Cramér - Wold principle, obtained independently by Cramér and Wold in 1932 using characteristic functions. Medical applications of Radon's result began in 1963 with the first tomography machine of A. Cormack and the use of the commercial tomography machine of G. Hounsfield. Both were awarded the Nobel prize in Medicine in 1979 for their work.

Computer tomography is based on the inversion of the Radon transformation, which allows us to reconstruct uniquely the density of a measure. The corresponding formulas were obtained by Radon. Other numerical algorithms can be found, for example, in Natterer. But such a unique reconstruction is possible only if one knows all, i.e. an infinite number of marginals. In practice one only has a finite number of marginals, or partially known data.
In some situations the incompleteness of the data is connected with the construction of the tomography machines, in others with the possibility of directing an $X-$ray through a part of the body. In both situations it is interesting to investigate how to reconstruct the density of the body and how precise the reconstruction is.

# Reconstruction of the Density from a Finite Number of Marginals

Let $Q_1$ and $Q_2$ be a pair of probabilities (measures) defined on the Borel $\sigma$-field of $\mathbb{R}^1$. Lorentz gave criteria for the existence of a probability density function on $\mathbb{R}^2$ taking only the values 0 or 1 and having $Q_1$ and $Q_2$ as marginals. Kellerer generalized this result, obtaining necessary and sufficient conditions for the existence of a density $f$ on $\mathbb{R}^2$ which satisfies the inequalities $0 \leq f(x) \leq 1$ and has $Q_1$ and $Q_2$ as marginals (see also Strassen and Jacobs. Lagarrias, Reeds and Shepp were able to show that Kellerer's and Lorentz's conditions are equivalent, i.e. for any density $f$ on $\mathbb{R}^2$ satisfying $0 \leq f(x) \leq 1$, there exists a density $g$ taking only the values 0 and 1 and having the same marginals. In general, similar results hold for probability densities on $\mathbb{R}^m$, $m \geq 2$, when the $(m-1)-$ dimensional marginals are prescribed.

Guttmann, Kempermann, Reeds and Shepp strengthened this result by showing that for any probability density$f$ on $\mathbb{R}^m$ satisfying $0 \leq f \leq 1$ and for any finite number of directions, there exists a probability density $g$ taking only the values 0, 1 and having the same marginals as $f$ in the chosen directions. It follows that densities having the same marginals in a finite number of arbitrary directions may differ considerably in the uniform metric. This leads to the following computer tomography paradox:

*For any human object and the corresponding projection data there exist many different reconstructions, in particular, a reconstruction, consisting only of bone and air (density 1 or 0), but still having the same projection data as the original object. Such non uniqueness results are common in tomography and are usually ignored because CT machines seem to produce useful images. It is likely that the "explanation" of this apparent paradox is that point reconstruction in tomography is impossible.*

Let us note that the existence of probability measures with fixed marginals is an important problem in the theory of probability metrics. This is especially true when studying the structure of minimal metrics (see Rachev). Most of our results will make use of relationships among different probability metrics, presented in a monographic form in Kakosyan, Klebanov and Rachev.
The purpose of this section is to show that under moment-type conditions, measures having a “large” number of coinciding marginals are close to each other in the weak metrics. Our method is based on techniques used in the classical moment problem. In showing that measures with large numbers of common marginals are close to each other in the weak metrics, the key idea is best understood by comparing three results. The first is the theorem of Guttmann, Kempermann, Reeds and Shepp mentioned above.

The second states that if a finite number of moments $\mu_1, \ldots, \mu_n$ of a function $f$ are given and $0 \leq f(x) \leq 1$, then there exists a function $g$ taking only the values 0 or 1 and having the same moments (see Karlin and Studden). It is clear that these two results are similar. However, the condition of equality of the marginals is more complex than the coincidence of the moments. Finally the third result gives estimates of the closeness in the $\lambda-$metric on $\mathbb{R}^1$ for measures having common moments $\mu_1, \ldots, \mu_n$ ($n < \infty$). These estimates are expressed in terms of the truncated Carleman series $\beta_m = \sum_{j=1}^{2m} \mu_{2j}^{-1/(2j)}$ ($2m < n$). The result shows that the closeness in the $\lambda-$metric is of order $\beta^{-1/4}$. Of course, since the condition of common marginals seems more restrictive than the condition of equal moments, one should be able to construct a similar estimate expressed in terms of the common marginals only. Furthermore, the technique required for such a construction should be similar to that used here.

Let us first derive estimates for closeness of measures in $\mathbb{R}^2$ having the same marginals in $n$ directions. We consider the case when one of the measures has a compact support, in which case the $\lambda-$closeness of measures has order $1/n$. Further, the compactness assumption will be relaxed by Carleman's assumption for the problem of moments. Here the $\lambda-$closeness of measures is of the order $\beta_{n/2}^{-1/4}$. We also derive estimates of the closeness of measures with the proportion $\epsilon$ $(0 < \epsilon < 1)$ of coinciding marginals. These estimates differ from the ones with equal marginals by an additional term of order $1/\log(1/\epsilon)$. We conclude by applying our results to the problems of computer tomography. In particular we offer a solution of the computer tomography paradox mentioned above.

To highlight the basic ideas, let us consider only the 2−dimensional case in full. Let $\theta_1, \ldots, \theta_n$ be $n$ unit vectors on the plane and let $\mathbb{P}_1, \mathbb{P}_2$ be two probability measures on $\mathbb{R}^2$, having the same marginals in the directions $\theta_1, \ldots, \theta_n$. Various probability metrics can be used to estimate the closeness of $\mathbb{P}_1$ and $\mathbb{P}_2$. We find it convenient to use the $\lambda-$ metric (see, e.g., Zolotarev), defined as

$$\lambda(\mathbb{P}_1, \mathbb{P}_2) = \min_{T>0} \max\{ \max_{\|t\| \le T} \mid \int_{\mathbb{R}^2} \exp(i\langle t, x\rangle)(\mathbb{P}_1 - \mathbb{P}_2)(dx) \mid, \frac{1}{T}\},$$

where $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ are the inner product and the Euclidean norm in $\mathbb{R}^2$, respectively. Clearly, the $\lambda$ metrizes the weak convergence.

In our first result, we consider the important case where one of the measures considered has a compact support.

### Theorem

*Let $\theta_1, \dots, \theta_n$ be $n \geq 2$ unit vectors in $\mathbb{R}^2$, no two of which are collinear. Let the support of the probability measure $\mathbb{P}_1$ be a subset of the unit disc, and let the probability measure $\mathbb{P}_2$ have the same marginals as $\mathbb{P}_1$ in the directions $\theta_1, \dots, \theta_n$. Then,*

$$\lambda(\mathbb{P}_1, \mathbb{P}_2) \leq \left(\frac{2}{s!}\right)^{1/(s+1)},$$

*where*

$$s = 2\left[\frac{n-1}{2}\right]$$

*and $[r]$ is the integer part of $r$.*

## Remark

*Note that we can replace the right-hand side of previous inequality by $C/s$, where $C$ is some constant, since $\left(\frac{2}{s!}\right)^{1/(s+1)} \sim \frac{e}{s}$ as $s \to \infty$.*

The above result leads to the following corollaries:

## Corollary

*Let $\theta_1, \ldots, \theta_n$ be $n \geq 2$ directions in $\mathbb{R}^2$ no two of which are collinear. Suppose that the marginals of the probabilities $\mathbb{P}_1$ and $\mathbb{P}_2$ with respect to the directions $\theta_1, \ldots, \theta_n$ have moments up to the even order $k \leq n-1$. Then the marginals of $\mathbb{P}_1$ and $\mathbb{P}_2$ with respect to any direction $t$ have the same moments up to order $k$.*

### Corollary

*Theorem 2 still holds if we replace the assumption that* $\mathbb{P}_1$ *and* $\mathbb{P}_2$ *have the same marginals with respect to the directions* $\theta_j\,(j = 1, \ldots, n)$ *with the assumption that these marginals have the same moments up to order* $n - 1$*.*

Let us now relax the condition of compactness for the support of $\mathbb{P}_1$, assuming only the existence of all moments together with the Carleman condition (which is a sufficient condition for the moments to determine the distribution uniquely, see, e.g., Harris (1966)). For convenience, let us introduce the following notation:

$$\mu_k = \sup_{\theta \in S^1} \int_{\mathbb{R}^2} \langle x, \theta \rangle^k \mathbb{P}_1(dx), \quad k = 0, 1, \ldots,$$

$$\beta_s = \sum_{j=1}^{(s-2)/2} \mu_{2j}^{-\frac{1}{2j}},$$

where the number $s$ is determined from (**??**) and $S^1$ is the unit circle.

### Theorem

*Let $\theta_1, \ldots, \theta_n$ be $n \geq 2$ directions in $\mathbb{R}^2$ no two of which are collinear. Suppose that the measure $\mathbb{P}_1$ has moments of any order. Suppose also that the marginals of the measures $\mathbb{P}_1$ and $\mathbb{P}_2$ in the directions $\theta_1, \ldots, \theta_n$ have the same moments up to order $n-1$. Then there exists an absolute constant $C$ such that*

$$\lambda(\mathbb{P}_1, \mathbb{P}_2) \leq C\beta_s^{-\frac{1}{4}}(\mu_0 + \mu_2^{1/2})^{1/4}.$$

### Theorem

*Suppose that, in addition to the conditions of previous Theorem, the characteristic function of the measure $\mathbb{P}_1$ admits analytic continuation in some disc centered at the origin. Then*

$$\lambda(\mathbb{P}_1, \mathbb{P}_2) \leq C_{\mathbb{P}_1} / \log(s),$$

*where the constant $C_{\mathbb{P}_1}$ depends on the measure $\mathbb{P}_1$, and not on the measure $\mathbb{P}_2$ or the number of directions.*

Let us now consider a more realistic situation where the marginals of $\mathbb{P}_1$ and $\mathbb{P}_2$ in the directions $\theta_1, \dots, \theta_n$ are not the same but are close in the metric $\lambda$. We use the same notation as that introduced in Theorem 2.

### Theorem

*Let $\theta_1, \dots, \theta_n$ be $n \geq 2$ directions in $\mathbb{R}^2$, no two of which are collinear. Suppose that the supports of the measures $\mathbb{P}_1$ and $\mathbb{P}_2$ lie in the unit disc, where they have $\epsilon-$coinciding marginals with respect to the directions $\theta_j$ $(j = 1, \dots, n)$ i.e.,*

$$\lambda(\mathbb{P}_1^{(\theta_j)}, \mathbb{P}_2^{(\theta_j)}) = \min_{T>0} \max \Big( \max_{|\tau| \leq T} \mid \varphi_1(\tau, \theta_j) - \varphi_2(\tau, \theta_j) \mid, 1/T \Big) \leq \epsilon,$$

*$j = 1, \dots, n$. Then there exists a constant $C$ depending on the directions $\theta_j$ $(j = 1, \dots, n)$ such that for sufficiently small $\epsilon > 0$, we have*

$$\lambda(\mathbb{P}_1, \mathbb{P}_2) \leq C(1/\log(\frac{1}{\epsilon}) + 1/s),$$

*where $s$ is defined as before.*

### Remark

*The conclusion of this Theorem still holds if instead of the $\epsilon-$ coincidence of the marginals we require the $\epsilon$-coincidence of the moments up to order $s$ of these marginals. For the latter, we require the inequalities*

$$|\int_{\mathbb{R}^2} \langle x, \theta_j \rangle^k \mathbb{P}_1(dx) - \int_{\mathbb{R}^2} \langle x, \theta_j \rangle^k \mathbb{P}_2(dx)| \leq \epsilon, \;\; k \leq s.$$

All the theorems stated above admit generalizations to probability measures defined on $\mathbb{R}^m$. However, here we can no longer choose the directions $\theta_1, \ldots, \theta_n$ in an arbitrary way. Furthermore, to obtain the order of precision corresponding to the $n$ directions in $\mathbb{R}^2$, we need $n^{m-1}$ directions in $\mathbb{R}^m$ for $m \geq 2$. The results can be obtained by induction on the dimension $m$.
We need to define the set of directions we are going to use. Choose $n \geq 2$ distinct real numbers $u_1, \ldots, u_n$, all different from zero, and construct first the set of $n$ two-dimensional vectors:

$$(1, u_1), (1, u_2), \ldots, (1, u_n).$$

Then, construct $n^2$ three-dimensional vectors

$$(1, u_{j_1}, u_{j_2}), \;\; j_1, j_2 = 1, \ldots, n.$$

Repeating the last step, we shall eventually construct $n^{m-1}$ vectors in $\mathbb{R}^m$:

$$(1, u_{j_1}, u_{j_2}, \ldots, u_{j_{m-1}}), \;\; j_l = 1, \ldots, n, \;\; l = 1, \ldots, m-1.$$

Denote these $m-$dimensional vectors by $\theta_1, \ldots, \theta_N$, where $N = n^{m-1}$ (the choice of enumeration is irrelevant here). These inductive arguments lead to the following extensions of previous Theorems.

## Proposition

*The results of Theorems 2 - 5 still hold if we consider the measures $\mathbb{P}_1$ and $\mathbb{P}_2$ in $\mathbb{R}^m$, and we choose as directions the $N = n^{m-1}$ vectors given above. Further, $s = 2\left[\frac{n-1}{2}\right]$.*

The above results are concerned with closeness between the probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ in terms of the $\lambda$-metric. We can also consider the cases of the Lévy-Prokhorov distance and the distance in variation with the additional assumptions of existence and differentiability of the densities of the relevant probability distributions. To obtain the corresponding estimates, it is sufficient to use the results relating these distances to the $\lambda$-metric. We do not formulate the corresponding theorems since the inequalities already obtained are far from being final. We belive the estimates of closeness between the densities of smoothed distributions are more interesting.

# Quantum mechanics and computer tomography

Let physical system has one continuous degree of freedom, having generalized coordinate operator $\hat{x}$ and conjugate operator $\hat{p}$. Let "rotated quadrature" operators $\hat{x}_\theta$, $\hat{p}_\theta$ are defined as

$$\hat{x}_\theta = \hat{x}\cos\theta + \hat{p}\sin\theta, \quad \hat{p}_\theta = -\hat{x}\sin\theta + \hat{p}\cos\theta$$

for all angles $\theta$ ($\hat{x}_\theta$, $\hat{p}_\theta$ are related with $\hat{x}$, $\hat{p}$ ($\theta = 0$) by unitary transformations). Let $P_\theta(x_\theta)$ is the probability distribution of observable $\hat{x}_\theta$. It was shown that if we know an infinite and continuous (uncountable) set of such distributions $\{P_\theta(x_\theta)\}$, then it is possible to determine in unique way by the Radon transform the corresponding Wigner distribution $W(x,p)$:

$$W(x,p) = \frac{1}{\pi}\int_{-\infty}^{\infty} \psi(x+x')\psi^*(x-x')e^{i2px'}dx'$$

which is in one-to-one relation to the wave function:

$$\psi(x+x')\psi^*(x-x') = \int_{-\infty}^{\infty} W(x,p)e^{i2px'}dp.$$

In Raymer, it was realized experimental method in QO for measuring of corresponding distributions $P_\theta(x_\theta; W)$

$$P_\theta(x_\theta; W) = \int_{-\infty}^{\infty} W(x_\theta \cos\theta -$$

$$-p_\theta \sin\theta, x_\theta \sin\theta + p_\theta \cos\theta) dp_\theta$$

which are marginals of $W(x, p)$. The Wigner distributions $W(x, p)$ cannot be, in general, measurable, because $W(x, p)$ are, in general, not non-negative. It was proved in Hudson that $W(x, p)$ are non-negative iff the wave functions are Gaussian.

So for "measuring" of the wave functions we must solve two inverse problems:

$$\{P_\theta(x_\theta; W)\} \to W(x,p) \quad (A);$$
$$W(x,p) \to \psi(x) \quad (B).$$

Both $(A)$ and $(B)$ inverse problems are ill-posed problems. Moreover, in reality we can measure only *finite* number $N$ of marginals of the Wigner function. But for finite number of marginals this non-uniqueness produces so called tomography paradox mentioned above. In our terms this means that two different Wigner distributions $W_1(x,p)$ and $W_2(x,p)$, which have the same $N$ marginals may differ dramatically. It is also an additional trouble - the Wigner distributions $W(x,p)$ cannot have finite support.

Therefore in CAT scans in reality are measured only a finite number of truncated marginals $\tilde{P}_\theta(x_\theta; W)$:

$$\tilde{P}_\theta(x_\theta; W) = \begin{cases} P_\theta(x_\theta; W), & x \in [X_{\theta_1}, X_{\theta_2}] \\ 0, & x \notin [X_{\theta_1}, X_{\theta_2}] \end{cases}$$

where *finite* $X_{\theta 1}, X_{\theta 2}$ determined by real possibilities of the experimental device. From the finiteness of the number of truncated marginals $\tilde{P}_\theta(x_\theta; W),\ \theta \in \{\theta_1, \ldots, \theta_N\}$, necessary discretization of the integral Radon transform (see Natterer) follows that the solution of problem $(A)$ has so-called artifacts.

Artifacts are non-existing details in the exact $W(x, p)$ and $\psi(x)$, which originate from the non-accuracy of mathematical part of the original CAT scans. Such artifacts easy to see in the results of the original works Raymer, Smeethy, even for the simplest case of the Gaussian $W(x, p)$.
We obtain main result.

## Theorem

*Let $\psi_1, \psi_2$ be two wave functions and $W_1, W_2$ be corresponding Wigner distributions. Suppose that $\psi_l$ $(l = 1, 2)$ is $2s$ time differentiable and*

$$|\psi_l(x)| \le C_1, \;\; \int_{-\infty}^{\infty} |\psi_l(x) dx| \le C_2,$$

$$\sup_a |\frac{\partial^{2s}}{\partial a^{2s}} e^{-\sigma^2 a^2/2} \int_{-\infty}^{\infty} \psi(u-\zeta)\psi^*(u) e^{iau\zeta} du| \le$$

$$\le C(|\zeta|^{2s} + 1),$$

$$g_l(a, \zeta; \sigma) =$$

$$= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_l(a - \zeta - \xi_1 - \xi_2) \psi_l^*(a - \xi_2) \times$$

$$\exp(-(\xi_1^2 + \xi_2^2)/(2\sigma^2)) d\xi_1 d\xi_2,$$

*$(l = 1, 2)$, where $C$, $C_1$, $C_2$ are positive constants. For any $\epsilon > 0$ and any integer $N \ge 2$ there exist $(N + 1)$ directions*

*$\theta_0, \dots, \theta_N$ such that if the marginals of $W_1$ and $W_2$ on these directions are $\epsilon$-identical, e.i.*

$$|P_{\theta_j}(x_{\theta_j}; W_1) - P_{\theta_j}(x_{\theta_j}; W_2)| \le \epsilon$$

*$j = 0, 1, \dots, N$, then for any $A > 0$*

$$\sup_a |g_1(a, \zeta; \sigma) - g_2(a, \zeta; \sigma)| \le$$

$$\frac{2}{\sigma}\Big[\frac{C\pi^{2s}A^s(1+\zeta^{2s})(8+\frac{4}{\pi}\ln N)}{\zeta^s 2^s (N+1)^2(N-1)^2\dots(N-2s+3)^2} +$$

$$+\epsilon\sqrt{\frac{\pi}{2}}(8+\frac{4}{\pi}\ln N) + C_1 C_2 \int_{\sigma A}^{\infty} e^{-x^2/2}dx\Big].$$

*The directions $\theta_0, \dots, \theta_N$ may be chosen such that the points*

$$a_0 = -1/\tan\theta_0, \dots, a_N = -1/\tan\theta_N$$

*are the Tchebyshev knots of interpolation for $[-A, A]$.*

From here we see that it is impossible to determine local information on the wave function basing on a finite number $N \geq 2$ of marginals, but it is possible to estimate the distance between two functions $g_1, g_2$, connected to the wave functions.