Some new advances in estimating the rate of convergence in Lyapunov's theorem

Ilya Tyurin, Moscow State University

The classical problem

$X_1, \dots, X_n$ are independent centered r.v. with finite variances.

Then

$$\rho(S_n, N) := \sup_x \left| \mathrm{P}(S_n \le x) - \mathrm{P}(N \le x) \right| \le C\varepsilon_n,$$

$$S_n = (X_1 + \dots + X_n) / \sigma(n), \quad \sigma(n)^2 = \sum_{k=1}^n \mathrm{E}X_k^2,$$

N is a standard normal r.v., $C = const$,

$$\varepsilon_n = \frac{\sum_{k=1}^n \mathrm{E}|X_k|^3}{\sigma(n)^3} -$$ Lyapunov's fraction.

Bounds for the constant C

	general case	*i.i.d. case*
Esseen, 1942:	$C \le 7{,}5$	
Bergstrom, 1949:	$C \le 4{,}8$	
Takano, 1951:		$C \le 2{,}031$
Zolotarev, 1965-1967:	$C \le 1{,}322$ $C \le 1{,}301$	
	$C \le 0{,}9051$	$C \le 0{,}8197$
Van Beek, 1972:	$C \le 0{,}7975$	
Shiganov, 1982:	$C \le 0{,}7915$	$C \le 0{,}7655$
Shevtsova, 2006:		$C \le 0{,}7056$
Tyurin, 2009:	$C \le 0{,}6379$	$C \le 0{,}5894$
	$C \le 0{,}5606$	$C \le 0{,}4785$
Korolev, Shevtsova, 2010:	$C \le 0{,}5600$	$C \le 0{,}4784$
Tyurin, 2010:	$C \le 0{,}5600$	$C \le 0{,}4784$
Tyurin, 2010:	$C \le 0{,}5591$	$C \le 0{,}4774$

It is known (Esseen, 1956), that $C \ge 0{,}409\ldots$

Ideal metrics ζ_r

For arbitrary $f \in C^{(r-1)}(\mathbb{R}),\ r \in \mathbb{N},$ define

$$M_r(f) := \sup_{x \neq y} \left| \frac{f^{(r-1)}(x) - f^{(r-1)}(y)}{x - y} \right|.$$

Let

$$\zeta_r(X,Y) := \sup\left\{ \left| \mathrm{E}f(X) - \mathrm{E}f(Y) \right| : f \in \mathcal{F}_r \right\}, \quad r = 1, 2, \ldots,$$

where $\mathcal{F}_r$ is the set of all real bounded functions with $M_r(f) \leq 1.$

ζ_1 has an important alternative representation

$$\zeta_1(X,Y) = \varkappa_1(X,Y) := \int_{-\infty}^{\infty} \left| \mathrm{P}(X \leq x) - \mathrm{P}(Y \leq x) \right| dx.$$

Ideal metrics ζ_r

Now suppose $r = k + s$, where k is integer, s – from (0,1]. Then

$$M_r(f) := \sup_{x \neq y} \frac{\left| f^{(k)}(x) - f^{(k)}(y) \right|}{\left| x - y \right|^s}.$$

As before,

$$\zeta_r(X,Y) := \sup\left\{ \left| \mathrm{E}f(X) - \mathrm{E}f(Y) \right| : f \in \mathcal{F}_r \right\}.$$

where $\mathcal{F}_r$ is the set of all real bounded functions with $M_r(f) \leq 1$.

If $0 < r < 1$, then ζ_r can be expressed as

$$\zeta_r(X,Y) = \inf\{ \mathrm{E}\left| \hat{X} - \hat{Y} \right|^r : Law(\hat{X}) = Law(X), Law(\hat{Y}) = Law(Y) \}$$

Ideal metrics

These metrics satisfy two conditions:

- regularity, i.e.

$$\zeta_r(X+Z, Y+Z) \le \zeta_r(X, Y)$$

for arbitrary *X*, *Y*, *Z* such that *Z* is independent of *X*, *Y*.

- homogeneity (of the order *r*), i.e.

$$\zeta_r(cX, cY) \le |c|^r \zeta_r(X, Y),$$

where *X*, *Y* are arbitrary r.v., *c* is a real number.

Zero bias transformation

Let W be centered r.v. with finite nonzero variance σ^2.

W^* has W-zero biased distribution if

$$\mathrm{E}Wf(W) = \sigma^2 \mathrm{E}f'(W^*)$$

for each real differentiable function f, such that l.h.s. is defined.

$Law(W) = Law(W^*)$ iff W is normal.

Zero bias transformation

Zero bias transformation was introduced by Goldstein and Reinert in 1997. It was used to estimate the convergence rate in CLT.

It was used in combination with Stein's method and allowed to reformulate the problem. Instead of

$$\sup\left\{\left|\mathrm{E}f(W)-\mathrm{E}f(N)\right| : f\in\mathcal{F}\right\}$$

one can consider

$$\sup\left\{\left|\mathrm{E}f(W)-\mathrm{E}f(W^*)\right| : f\in\mathcal{F}^*\right\}.$$

Zero bias transformation of a normed sum

Let $X_1, \dots, X_n$ be i.i.d. r.v. with E$X_1 = 0$, Var $X_1 = 1$. X_1^* is independent from $X_1, \dots, X_n$ and has X_1-zero biased distribution.

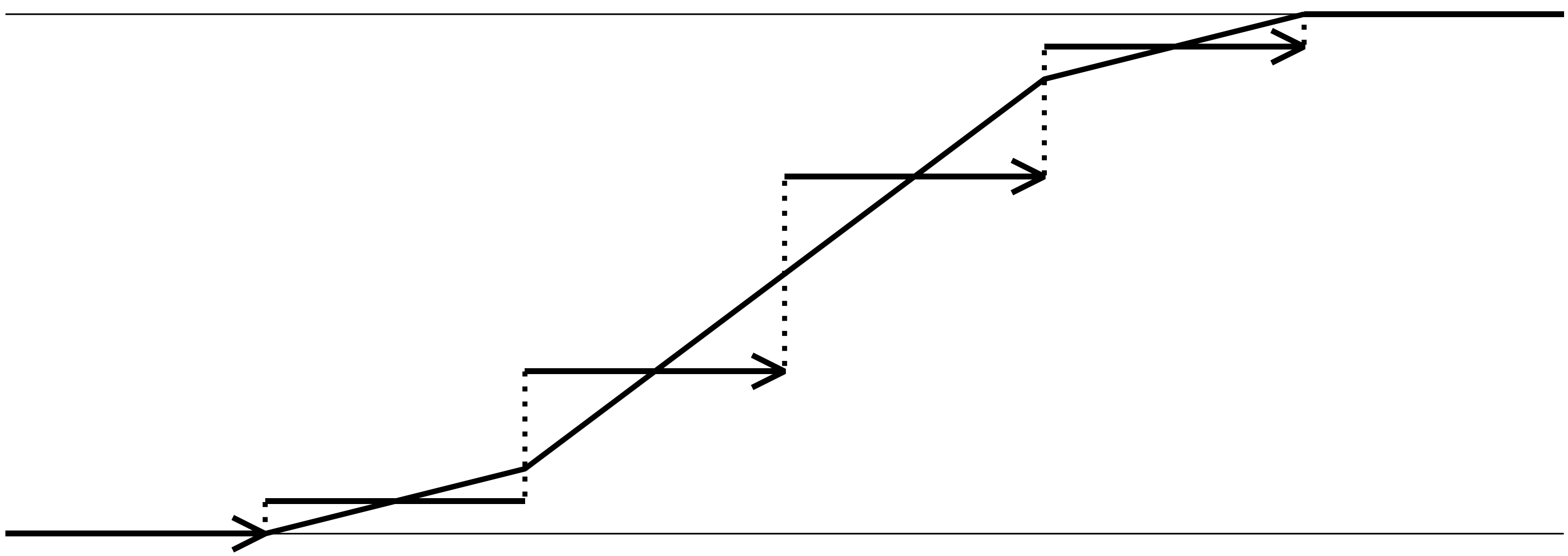
Then

$$\left(\frac{X_1 + \ldots + X_n}{\sqrt{n}}\right)^* \overset{Law}{=} \frac{X_1^* + \ldots + X_n}{\sqrt{n}}.$$

As a result one has

$$\zeta_1\left(S_n, S_n^*\right) \leq \zeta_1\left(\frac{X_1}{\sqrt{n}}, \frac{X_1^*}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}}\zeta_1\left(X_1, X_1^*\right).$$

Zero bias transformation of a normed sum



Example: distribution functions of S_n and S_n*

Stein's method

W – centered r.v., $\mathrm{Var}W = 1$, N – standard normal r.v., f – real function.

$$h(w) := e^{\frac{w^2}{2}} \int_{-\infty}^{w} (f(w) - \mathrm{E}f(N)) e^{-\frac{x^2}{2}} dx$$

satisfies Stein's equation

$$h'(w) - wh(w) = f(w) - \mathrm{E}f(N).$$

Putting $w=W$ and taking expectations from both sides yields

$$|\mathrm{E}f(W) - \mathrm{E}f(N)| = |\mathrm{E}h'(W) - \mathrm{E}Wh(W)| = |\mathrm{E}h'(W) - \mathrm{E}h'(W^*)| \leq$$
$$\leq M_2(h)\zeta_1(W, W^*).$$

$$M_2(h) \leq \left\{ 2M_1(f), \frac{\sqrt{2\pi}}{4} M_2(f), \frac{1}{3} M_3(f) \right\} \qquad \text{(Raic, 2003).}$$

Stein's method

$$|\mathrm{E}f(W)-\mathrm{E}f(N)|\le M_2(h)\zeta_1(W,W^*),$$

$$M_2(h)\le\left\{2M_1(f),\frac{\sqrt{2\pi}}{4}M_2(f),\frac{1}{3}M_3(f)\right\}$$

$$\Downarrow$$

$$\zeta_1(W,N)\le 2\zeta_1(W,W^*),\quad \zeta_2(W,N)\le\frac{\sqrt{2\pi}}{4}\zeta_1(W,W^*),$$

$$\zeta_3(W,N)\le\frac{1}{3}\zeta_1(W,W^*).$$

Theorem

Let W be a centered r.v. with $\mathrm{Var}W = 1$, then

$$\zeta_1(W, W^*) \le \frac{\mathrm{E}|W|^3}{2},$$

with equality when W has a 2-point distribution.

Moreover, for S_n and ε_n as defined before

$$\zeta_1(S_n, S_n^*) \le \frac{1}{2}\varepsilon_n.$$

Quasiconvex functions

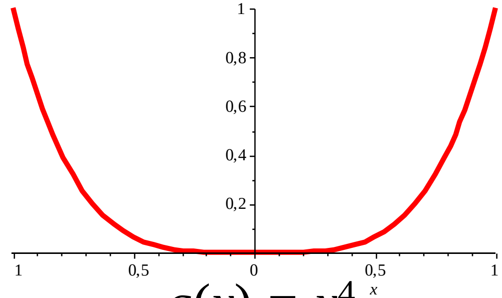
G – convex subset of a linear space

$g\colon G \to \mathbb{R}$ – convex

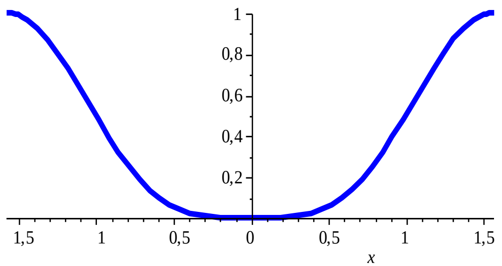
$$g(\alpha v_1 + (1-\alpha)v_2) \le \alpha g(v_1) + (1-\alpha)g(v_2)$$

g – quasiconvex

$$g(\alpha v_1 + (1-\alpha)v_2) \le \max\{g(v_1), g(v_2)\}$$



$g(x) = x^4$



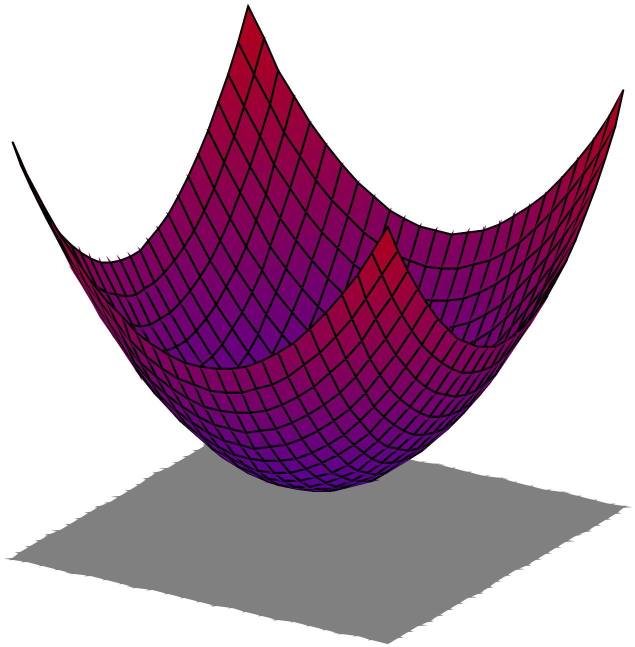
$g(x) = \sin(x)^4$

Quasiconvex functions

A quasiconvex function defined on a polytope attains its supremum at one of the vertices

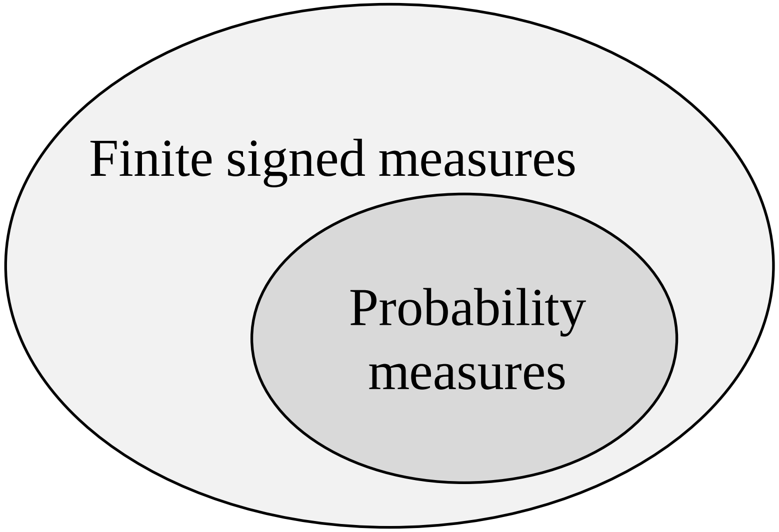
<u>Example</u>

$f(x,y)=x^2+y^2$ defined on a rectangle $[-5,5]\times[-5,5]$



$f(x,y)=x^2+y^2$

Quasiconvex functions



(S,ρ) – metric space

Q = {all finite signed measures on $B(S)$}
D = {discrete probability measures on $B(S)$}

Q with traditional operations is a linear space and D is a convex subset of Q.

$$\mathrm{E}f(X_1,\ldots,X_n) = \int_{\mathbb{R}^n} f\,d\mathrm{P}_{X_1}\ldots d\mathrm{P}_{X_n}$$

f: $\mathbb{R} \to \mathbb{R}^n$, $\mathrm{P}_{X_1}, \ldots, \mathrm{P}_{X_n}$ – distributions of r.v. $X_1, \ldots, X_n$.

Reduction theorem

$h_1, \dots, h_m$ – real functions on S

$$K := \{\mu \in D, \langle h_i, \mu \rangle = 0, i = 1,\dots,m\} \qquad \langle f, \mu \rangle := \int_S f d\mu$$

D_j = {j-point distributions}, $\quad D_{\le j} = D_1 \bigcup \dots \bigcup D_j,$

$$K_j := \{\mu \in K \bigcap D_{\le j}\}, \ j = 1,2,\dots$$

For every quasiconvex function $g: K \to \mathbb{R}$

$$\sup_{\mu \in K} g(\mu) = \sup_{\mu \in K_{m+1}} g(\mu).$$

Extremal property of 3-point distributions

Take $h_1(x) = x$, $h_2(x) = x^2 - 1$. Then K is the set of discrete distributions corresponding to a centered r.v. with unit variance.

$$\zeta_1(W,W^*) = \int_{x=-\infty}^{\infty} \left| \mathrm{P}_W\big((-\infty,x]\big) - \mathrm{P}_{W^*}\big((-\infty,x]\big) \right| dx$$

is a convex function of P_W.

$$\mathrm{E}|W|^3 = \int_{\mathbb{R}} |x|\, \mathrm{P}_W(dx)$$

is linear w.r.t. P_W. Nonnegative convex function / positive linear function = = quasiconvex. Hence $\zeta_1(W,W^*)/\mathrm{E}|W|^3$, when considered as a function of the distribution P_W is quasiconvex.

$\sup\{\zeta_1(W,W^*)/\mathrm{E}|W|^3 : \mathrm{E}W = 0, \mathrm{Var}W = 1\}$ can be obtained by examining 2- and 3-point distributions.

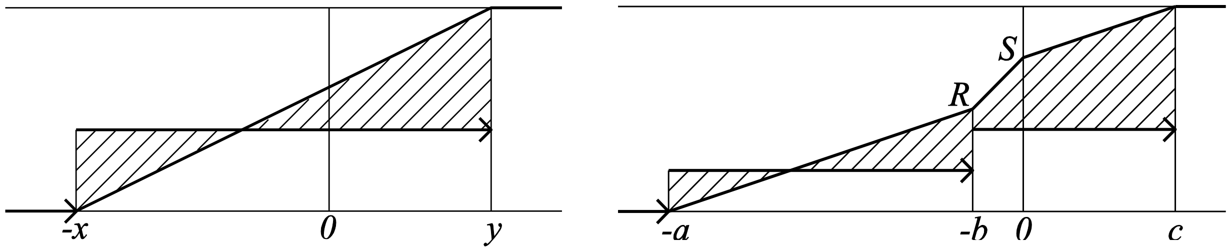
Extremal property of 2-point distributions

It turned out that for every centered r.v. W with $\mathrm{Var}W = 1$ that takes 3 different values holds the inequality

$$\zeta_1(W, W^*) \le \frac{1}{2}\mathrm{E}|W|^3.$$

For r.v. W that takes 2 different values one has

$$\zeta_1(W, W^*) = \frac{1}{2}\mathrm{E}|W|^3.$$



Estimates in metrics ζ_r

$$\zeta_1(W,N) \le 2\zeta_1(W,W^*), \quad \zeta_2(W,N) \le \frac{\sqrt{2\pi}}{4}\zeta_1(W,W^*),$$

$$\zeta_3(W,N) \le \frac{1}{3}\zeta_1(W,W^*).$$

$$\Downarrow \qquad \zeta_1(S_n,S_n^*) \le \frac{1}{2}\varepsilon_n$$

$$\zeta_1(S_n,N) \le \varepsilon, \quad \zeta_2(S_n,N) \le \frac{\sqrt{2\pi}}{8}\varepsilon,$$

$$\zeta_3(S_n,N) \le \frac{1}{6}\varepsilon.$$

Theorem

Let W be a centered r.v. with $\mathrm{Var}W = 1$. Let $d(x)$ be a nondecreasing concave function on $x \geq 0$ with $d(+0) = 0$.

Define

$$\tau_d(X,Y) = \inf\left\{\mathrm{E}d\left(|\hat{X} - \hat{Y}|\right) : Law(\hat{X}) = Law(X), Law(\hat{Y}) = Law(Y)\right\}$$

One has

$$\tau_d(W, W^*) \leq \mathrm{E}d\left(\left|W^*\right|\right).$$

In particular, for $0 < r < 1$

$$\zeta_r(W, W^*) \leq \frac{\mathrm{E}|W|^{2+r}}{1+r}.$$

Estimates in metrics $\zeta_{2+\delta}$

Suppose now, that $E|X_1|^{2+\delta}$ is finite for some $0 < \delta < 1$.

Stein's method in combination with the zero bias technique allows to establish the estimate

$$\zeta_{2+\delta}(S_n, N) \le \frac{E|X_1|^{2+\delta}}{(1+\delta)(2+\delta)n^{\delta/2}}.$$

Zero biasing and characteristic functions

Theorem

Let $f(s)$ be the characteristic function of centered r.v. W with unit variance, $f^*(s)$ – characteristic function of W^*. Then

$$| f(t) - \varphi(t) | \leq \varphi(t) \int_0^t | f(s) - f^*(s) | \, s \exp\left(\frac{s^2}{2} \right) ds, \quad t \in \mathbb{R}.$$

Here $\varphi(t) = \exp(-t^2/2)$.

Remark

$$f'(t) = \mathrm{E} iW e^{itW} = i\mathrm{E}W\cos(tW) - \mathrm{E}W\sin(tW) =$$

$$= -it\mathrm{E}\sin(tW^*) - t\mathrm{E}\cos(tW^*) = -tf^*(t).$$

Zero biasing and characteristic functions

Let $X_1, \dots, X_n$ be i.i.d. r.v. with E$X_1 = 0$, Var $X_1 = 1$. X_1^* is independent from $X_1, \dots, X_n$ and has X_1-zero biased distribution.

Then

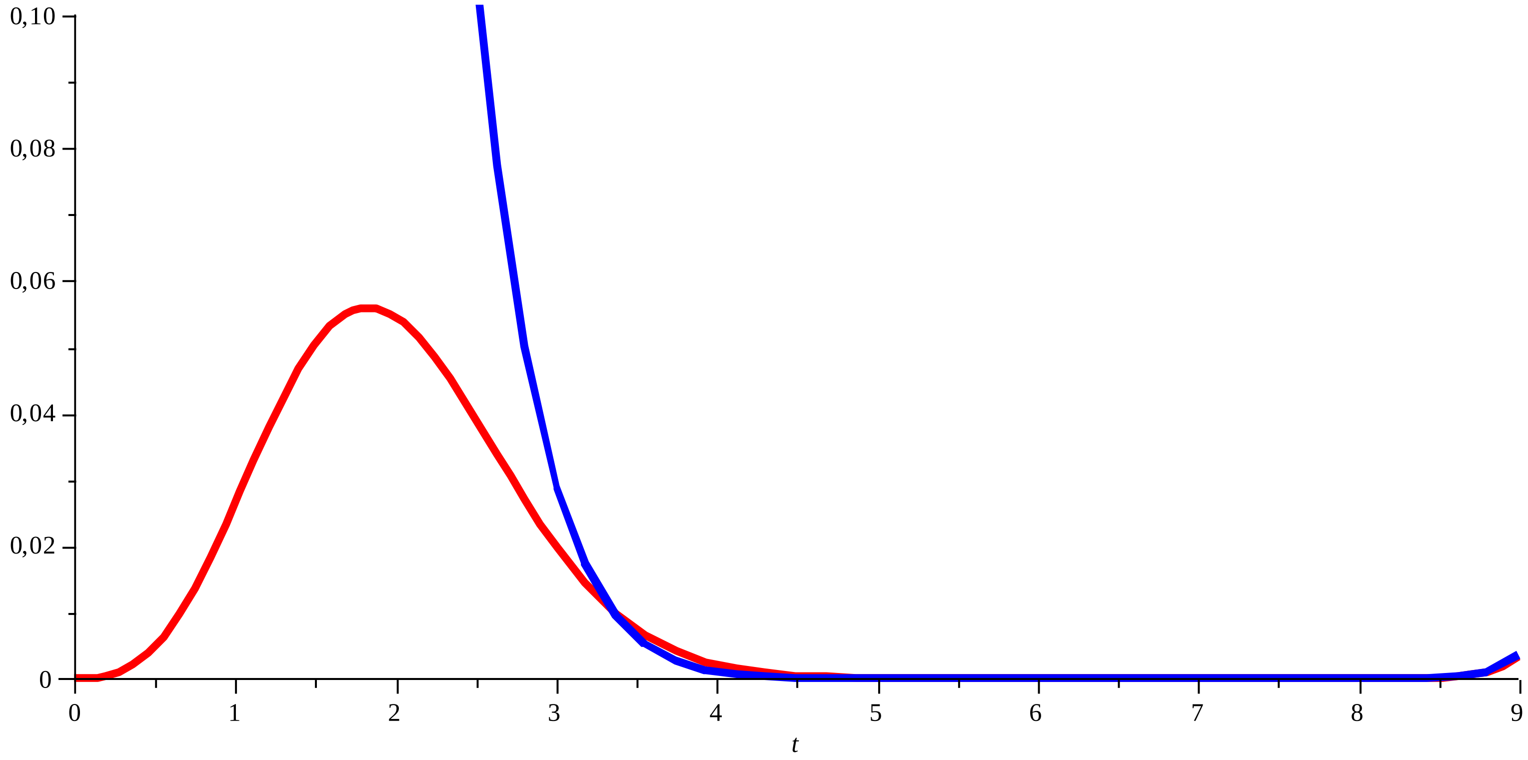
$$\left(\frac{X_1 + \ldots + X_n}{\sqrt{n}}\right)^* \overset{Law}{=} \frac{X_1^* + \ldots + X_n}{\sqrt{n}}.$$

Denote $f_1(t)$ and $f_1^*(t)$ respectively characteristic functions of X_1 and X_1^*. Then

$$f_{S_n}(t) - f_{S_n}^*(t) = f_1\left(\frac{t}{\sqrt{n}}\right)^{n-1}\left(f_1\left(\frac{t}{\sqrt{n}}\right) - f_1^*\left(\frac{t}{\sqrt{n}}\right)\right),$$

$$\left|f_1(t) - f_1^*(t)\right| \le |t|\,\zeta_1(X_1, X_1^*).$$

Zero biasing and characteristic functions



estimate for $\delta_{16}(t)$, $\varepsilon = 0{,}25$

Conclusion

- Optimal estimates for the proximity between a distribution and its zero bias transformation.
- Sharp estimates of the convergence rate in terms of ideal metrics ζ_r, where $1 \le r \le 3$.
- Sharp estimates for the proximity of characteristic functions.
- New estimates allowed to prove that in general case $C \le 0.5591$. And in the case of i.i.d. r.v. $C \le 0.4774$.