CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

Convolution Metrics
and
Rates of Convergence in the CLT
by
S.T. Rachev
and
J.E. Yukich

Technical Report No. 210
October 1987
Convolution metrics and rates of convergence in the CLT (Central Limit Theorem)

Let $B$ be a complete separable Banach space and let $V$ be its vector space of all random variables defined in a probability space $(\Omega, \mathcal{F}, P)$ and taking values in $B$. It is known that metrics on $B$ of convolution type enjoy a variety of interesting properties. In this article, it is shown that convolution metrics can also be used to obtain rates of convergence in CLT's involving a stable limit law. The rates are expressed in terms of a variety of uniform metrics on $B$ and include the total variation metric and the uniform metrics between density and characteristic functions. The results represent both an improvement and an extension of existing results. Weak convergence properties of convolution metrics are also explored.
Convolution Metrics
and
Rates of Convergence in the CLT

by

S. T. Rachev*1
Center for Stochastic Processes
Department of Statistics
University of North Carolina

and

J.E. Yukich
Institute of Mathematics
Sofia, Bulgaria

*On leave from Institute of Mathematics, Sofia, Bulgaria.

Research supported by the Air Force Office of Scientific Research Grant
No. F49620 85C 0144.
Abstract. Let \((B, \|\cdot\|)\) be a complete separable Banach space and let 
\(\mathcal{X} = \mathcal{X}(B)\) be the vector space of all random variables defined on a 
probability space \((\Omega, \mathcal{F}, P)\) and taking values in \(B\). It is known that 
metrics on \(\mathcal{X}\) of convolution type enjoy a variety of interesting 
properties. In this article it is shown that convolution metrics may 
also be used to obtain rates of convergence in CLT's involving a stable 
limit law. The rates are expressed in terms of a variety of uniform 
metrics on \(\mathcal{X}\) and include the total variation metric and the uniform 
metrics between density and characteristic functions. The results 
represent both an improvement and an extension of existing results. 
Weak convergence properties of convolution metrics are also explored.

AMS 1980 Subject Classifications
Primary: 60F05, 60G50, 60B10 Secondary: 60B12, 60E07

Key words and phrases: ideal probability metrics, convolution metrics, 
rates of convergence, stable random variables.
§0. Introduction

In this article we shall be concerned with various convolution pseudo-metrics on $\mathcal{P}(\mathbb{R}^k)$, or more generally $\mathcal{P}(B)$, $B$ a Banach space, where $\mathcal{P}(B)$ denotes the collection of all Borel probability measures on $B$. By convolution pseudo-metrics on $\mathcal{P}(\mathbb{R}^k)$ we mean distances of the form

$$d_g(P,Q) = \sup_{x \in \mathbb{R}^k} \left| \int g(x-y)(dP-dQ)(y) \right| \quad P, Q \in \mathcal{P}(\mathbb{R}^k),$$

where $g$ is a kernel belonging to $L^1(\mathbb{R}^k) \cap C_0(\mathbb{R}^k)$ and where $C_0(\mathbb{R}^k)$ is the collection of bounded continuous functions on $\mathbb{R}^k$ vanishing at infinity. Every such kernel generates an associated distance $d_g$ on $\mathcal{P}(\mathbb{R}^k)$. When the set $\{t : \hat{g}(t) = 0\}$ has empty interior, $d_g$ is actually a metric (here $\hat{g}$ denotes the Fourier transform of $g$); we refer the reader to Theorem A below for more on this.

We recall that these convolution pseudo-metrics enjoy a variety of properties and that they are actually quite useful from a statistical point of view. For example, as the following shows, one may characterize those convolution pseudo-metrics metrizing weak convergence and one can also obtain a CLT for $d(P_n, P)$, where here and henceforth $P_n$ denotes the $n^{th}$ empirical probability measure for $P$.

**Theorem A** [Yukich (1985)] Let $g \in L^1(\mathbb{R}^k) \cap C_0(\mathbb{R}^k)$. The following are equivalent statements for the pseudo-metric $d_g$:

(i) $d_g$ metrizes the topology of weak convergence in $\mathcal{P}(\mathbb{R}^k)$,

(ii) the set $\{t : \hat{g}(t) = 0\}$ has empty interior, and

(iii) $d_g(P,Q) \neq 0$ if $P \neq Q.$
\textbf{Theorem B} [Yukich (1985)] Let $g \in L^1(\mathbb{R}^k) \cap C^1(\mathbb{R}^k)$ be a decreasing function of $\| \cdot \|$. Then for all $P \in \mathcal{F}(\mathbb{R}^k)$ we have

$$n^{1/2} \, d_n(P, P) \overset{D}{\to} \sup_{x \in \mathbb{R}^k} \left| G_p(g(x-)) \right|,$$

where $G_p$ is a mean zero Gaussian process indexed by the translates of $g$.

Of course, by drawing on the results of Dudley and Philipp (1983), one could easily deduce a bounded LIL result from (0.1), but we will not pursue this here. Likewise, from the theory of the function indexed empirical process, one can also obtain refined exponential bounds for $n^{1/2} \, d_n(P, P)$. Also, the above results can be easily extended to the group setting; see Yukich (1987).

Keeping in mind that we are free to choose from among a variety of kernels $g$, it is sometimes relatively easy to calculate $d_n(P, Q)$. From a statistical viewpoint, this is important. For example, if $P$ is a uniform probability measure then $d_n(P, P)$ may be easily evaluated if $g$ is the bilateral exponential kernel. For more on this, see Yukich (1987). In this context we should mention that other well known metrics, e.g. the Prokhorov and dual bounded Lipschitz metrics, are relatively difficult to calculate and moreover, do not always satisfy the general CLT result (0.1); see the recent results of Gine and Zinn (1987).

This article, which may be regarded as an extension of previous work surrounding convolution metrics, shows that certain modified convolution metrics are useful in another context: they are extremely appropriate for providing rates of convergence (with respect to a variety of metrics) in the general CLT. In fact, some of our results will hold in the Banach space setting. We will also examine the close relationship between convolution metrics and the Kantorovich-Wasserstein distance. Thus, convolution metrics enjoy a variety of useful properties, making their possible statistical use seem...
especially attractive.

Let us be more precise about the contents of this article. Let (B, ||·||) be a complete separable Banach space equipped with the usual Borel sets \( \mathcal{B} \) and let \( \mathcal{X} := \mathcal{X}(B) \) be the vector space of all random variable defined on a probability space \( (\Omega, \mathcal{A}, \mathbb{P}) \) and taking values in B. We will choose to work with metrics on the space \( \mathcal{X} \) instead of the space \( \mathcal{F}(B) \). We will show that certain convolution metrics on \( \mathcal{X} \) may be used to provide improved rates of convergence of normalized sums to a stable limit law. The rates of convergence, which hold uniformly in \( n \), are expressed in terms of a variety of uniform metrics on \( \mathcal{X} \). In the Banach space setting the convergence rates hold with respect to the total variation metric and represent both an improvement and an extension of existing results. Even in the classical Euclidean space setting our approach provides improved rates of convergence and it also allows a determination of convergence rates with respect to uniform metrics between density and characteristic functions.

§1. Definitions, notation and terminology

A mapping \( \mu: \mathcal{X} \times \mathcal{X} \to [0, \infty] \) is called an ideal probability metric of order \( r \in \mathbb{R} \) if for any random variables \( X_1, X_2, Z \in \mathcal{X} \) and any non-zero constant \( c \) the following two properties are satisfied [cf. Zolotarev(1976)]:

(i) Regularity: \( \mu(X_1 + Z, X_2 + Z) \leq \mu(X_1, X_2) \), and

(ii) Homogeneity of order \( r \):

\[
\mu(cX_1, cX_2) = |c|^r \mu(X_1, X_2).
\]

When \( \mu \) is a simple metric, i.e., its values are determined by the marginal distributions of the random variables being compared, then it is assumed in addition that the random variable \( Z \) is independent of \( X_1 \) and \( X_2 \) in condition (i). All metrics \( \mu \) in this article are simple.

Zolotarev (1976) showed the existence of an ideal metric of a given
order $r \geq 0$ and he defined the ideal metric

$$
\zeta_r(X_1, X_2) := \sup \{ |E(f(X_1) - f(X_2))| : |f^{(m)}(x) - f^{(m)}(y)| \leq ||x - y||^\beta \},
$$

where $m \in \mathbb{N}^+$ and $\beta \in (0, 1]$ satisfy $m + \beta = r$ and $f^{(m)}$ denotes the $m$th Fréchet derivative of $f$ for $m \geq 0$ and $f^{(0)}(x) = f(x)$. He also obtained an upper bound for $\zeta_k$, $k \in \mathbb{N}^+$, in terms of the so-called difference pseudomoment $\kappa_k$, where for $r > 0$

$$
\kappa_r(X_1, X_2) := \sup \{|E(f(X_1) - f(X_2))| : |f(x) - f(y)| \leq ||x||^r \cdot ||y||^r \cdot ||x - y||^r \}.
$$

If $B = \mathbb{R}$, $||x|| = |x|$, then

$$
\kappa_r(X_1, X_2) := r \int_{-\infty}^{\infty} |x|^{r-1} |F_{X_1} - F_{X_2}| dx, \quad r > 0,
$$

and where $F_X$ denotes the distribution function for $X$.

In this article we introduce and study two ideal metrics of convolution type on the space $\mathcal{X}$. In addition to their ideality and convolution structure, these metrics have the following useful and special properties:

(P1) they have upper bounds which can be explicitly calculated in terms of the so-called difference pseudomoments, and

(P2) they have a weaker uniform structure than most other ideal metrics and thus yield better rates of convergence. These ideal metrics will be used to provide improved convergence rates for convergence to an $\alpha$-stable random variable in the Banach space setting. Moreover, the rates will hold with respect to a variety of uniform metrics on $\mathcal{X}$. 
More precisely, letting $X, X_1, X_2, \ldots$ denote i.i.d. random variables and $Y_\alpha$ denote an $\alpha$-stable random variable we use ideal metrics to describe the rate of convergence

\begin{equation}
\frac{X_1 + \ldots + X_n}{n^{1/\alpha}} \overset{D}{\to} Y_\alpha
\end{equation}

with respect to the following uniform metrics on $\mathfrak{X}$.

**Distance in variation metric:**

\begin{equation}
\sigma(X_1, X_2) := \sup_{A \in \mathfrak{B}} \left| P(X_1 \in A) - P(X_2 \in A) \right|, \quad X_1, X_2 \in \mathfrak{X}(\mathbb{B}),
\end{equation}

\begin{equation}
:= \sup \{|\mathbb{E}f(X_1) - \mathbb{E}f(X_2)| : f: \mathbb{B} \to \mathbb{R} \text{ is measurable and } \forall x, y \in \mathbb{B} \mid f(x) - f(y) \mid \leq \Upsilon(x, y) \text{ where } \Upsilon(x, y) = 1 \text{ if } x \neq y \text{ and } 0 \text{ otherwise}\},
\end{equation}

**Total variation metric:**

\begin{equation}
\text{Var}(X_1, X_2) := \mathbb{L}_1(X_1, X_2) := \sup \{|\mathbb{E}f(X_1) - \mathbb{E}f(X_2)| : f: \mathbb{B} \to \mathbb{R} \text{ is measurable and } ||f||_{\infty} := \sup_{x \in \mathbb{B}} |f(x)| \leq 1 \}
\end{equation}

\begin{equation}
= 2\sigma(X_1, X_2), \quad X_1, X_2 \in \mathfrak{X}(\mathbb{B}).
\end{equation}

In $\mathfrak{X}(\mathbb{R}^n)$ we have

\begin{equation}
\text{Var}(X_1, X_2) := \int |d(F_{X_1} - F_{X_2})|.
\end{equation}

**Uniform metric between densities ($p_X$ denotes the density for $X \in \mathfrak{X}(\mathbb{R}^k)$):**

\begin{equation}
\mathbb{L}(X_1, X_2) := \text{esssup}_{x} |p_{X_1}(x) - p_{X_2}(x)|.
\end{equation}

**Uniform metric between characteristic functions:**
\[
\chi(X_1, X_2) := \sup_t |\varphi_{X_1}(t) - \varphi_{X_2}(t)|,
\]

where \(\varphi_X\) denotes the characteristic function of \(X\). The metric \(\chi\) is topologically weaker than \(\text{Var}\), which is itself topologically weaker than \(\mathcal{L}\) by Scheffé's Theorem; see Billingsley (1968), p. 224.

The convergence rates with respect to \(\sigma\) will hold in the Banach space setting, thereby extending and generalizing results of Zolotarev (1976, 1977), Senatov (1980) and Paulauskas (1973, 1976) who consider uniform rates in terms of \(\xi_r\) and \(\kappa_r\) as well as stronger metrics.

Even in the special case \(B = \mathbb{R}\), our results improve upon those of Senatov (1980) and Paulauskas (1973, 1976), primarily because of the special property (P2). Our results with respect to the \(\chi\) and \(\mathcal{L}\) metrics, which seem to be the first of their kind, will follow from the easy applicability of the methods used for the \(\text{Var}\) metric. We note that not only does our method enjoy wide applicability to a variety of situations, but it is also remarkably simple.

Following this brief introductory remark, let us describe the contents of this article. This section concludes with notational remarks and section two discusses ideal convolution metrics and their properties, especially (P1) and (P2).

In section three we use an ideal convolution metric to obtain the rate of convergence in (1.3) in terms of the metric \(\text{Var}\) in the Banach space setting. Sections four and five illustrate the wide applicability of our method and describe rates in terms of the uniform metrics \(\chi\) and \(\mathcal{L}\), respectively. Here, the unifying theme is that ideal metrics, especially those of convolution type, provide improved convergence rates in (1.3) with respect to various uniform metrics, e.g., \(\text{Var}\), \(\chi\), or \(\mathcal{L}\).

Section six shows that the new ideal convolution metrics metrize the convergence in distribution of random variables and section seven contains concluding remarks.

**Notation and Terminology**

For each \(X_1, X_2 \in \mathcal{X}\) we write \(X_1 + X_2\) to mean the sum of
independent random variables with laws \( P_{X_1} \) and \( P_{X_2} \), respectively. For any \( X \in \mathbb{X} \), \( p_X \) denotes the density of \( X \) if it exists. We reserve the letter \( Y_\alpha \) (or \( Y \)) to denote a strictly stable symmetrical random variable with parameter \( \alpha \in (0,2) \), i.e., \( Y_\alpha \overset{D}{=} -Y_\alpha \) and for any \( n = 1,2, \ldots \), \( X_1' + \ldots + X_n' \overset{D}{=} n^{1/2} Y_\alpha \), where \( X_1', X_2', \ldots, X_n' \) are i.i.d. random variables with the same distribution as \( Y_\alpha \). If \( Y_\alpha \in \mathbb{X}(\mathbb{R}) \) we assume that \( Y_\alpha \) has characteristic function

\[
\varphi_\alpha(t) = \exp(-|t|^\alpha), \quad t \in \mathbb{R}.
\]

For any \( f : \mathbb{B} \to \mathbb{R} \), \( \|f\|_L := \sup_{x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|} \) denotes the Lipschitz norm of \( f \), \( \|f\|_\infty \) the essential supremum of \( f \), and when \( \mathbb{B} = \mathbb{R}^k \), \( \|f\|_p \) denotes the \( L^p \) norm.

We will use the following metrics on \( \mathbb{X} \):

Kolmogorov metric:

\[
\rho(X_1, X_2) := \sup_{x \in \mathbb{R}} |F_{X_1}(x) - F_{X_2}(x)|, \quad X_1, X_2 \in \mathbb{X}(\mathbb{R}).
\]

Weighted \( \chi \) metric:

\[
\chi_\tau(X_1, X_2) := \sup_{t \in \mathbb{R}} |t|^{-\tau} |\varphi_{X_1}(t) - \varphi_{X_2}(t)|,
\]

\( L^p \) version of \( \zeta_m \):

\[
\zeta_{m,p}(X_1, X_2) := \sup_{|f(X_1) - f(X_2)| : \|f^{(m+1)}\|_q \leq 1} \frac{1}{p} + \frac{1}{q} = 1, \quad m = 0, 1, 2, \ldots
\]
(If $\zeta_{m,p}(X_1,X_2) \leq \infty$ then \(\zeta_{m,p}(X_1,X_2) = \|\int_{-\infty}^{*} \frac{(s-t)^m}{m!} d(F_{X_1}(t)-F_{X_2}(t))\|_p\).

Generalized Kantorovich-Wasserstein metric:

\[
W_p(X_1,X_2) := \sup \{ \int f dF_{X_1} + \int g dF_{X_2} : \|f\|_\infty + \|f\|_L < \infty, \\
\|g\|_\infty + \|g\|_L < \infty \text{ and } f(x) + g(y) \leq \|x-y\|^p \quad \forall x,y \in B \}, \quad p \geq 1.
\]

§2. Ideal convolution metrics and their properties

Let \(\theta \in \mathcal{X}(\mathbb{R}^k)\) and define for every \(r > 0\) the convolution metric

\[
\mu_{\theta,r}(X_1,X_2) := \sup_{h \in \mathbb{R}} \|h\|^r \ell(X_1+h\theta,X_2+h\theta) \quad X_1,X_2 \in \mathcal{X}(\mathbb{R}^k).
\]

Thus, each random variable \(\theta\) generates a metric \(\mu_{\theta,r}, r > 0\). When \(\theta \in \mathcal{X}(\mathbb{B})\) we will also consider convolution metrics of the form [cf. Rachev and Ignatov (1984)]:

\[
\nu_{\theta,r}(X_1,X_2) := \sup_{h \in \mathbb{R}} \|h\|^r \ell(X_1+h\theta,X_2+h\theta) \quad X_1,X_2 \in \mathcal{X}(\mathbb{B}).
\]

Lemmas 2.1 and 2.2 below show that \(\mu_{\theta,r}\) and \(\nu_{\theta,r}\) are ideal of order \(r-1\) and \(r\), respectively. In general, \(\mu_{\theta,r}\) and \(\nu_{\theta,r}\) are actually only pseudo-metrics (cf. Theorem A), but this distinction is not of importance in what follows and so we omit it.

When \(\theta\) is an \(\alpha\)-stable random variable we will write \(\mu_{\alpha,r}\) and \(\nu_{\alpha,r}\) (or simply \(\mu_r\) and \(\nu_r\) when it is understood) in place of \(\mu_{\theta,r}\) and \(\nu_{\theta,r}\). Also, if \(\theta\) has a density \(g\), then the metric \(\mu_{\theta,r}\) represents a generalization of the convolution metric
d_\mathcal{g}(X_1, X_2) := \ell(X_1 + h\theta, X_2 + h\theta)

described in the introduction.

One of the central themes of this article is that every ideal metric on \( \mathcal{E} \) can be used to provide convergence rates in (1.3) in terms of some uniform metric (e.g. \( \text{Var} \), \( \chi \) or \( \ell \)) corresponding to the proposed ideal metric.

For example, we show that \( \mu_r \) can be used to describe the convergence rate in (1.3) with respect to the uniform metric \( \text{Var} \). The method for the ideal metric \( \mu_r \) is strikingly simple and yet general enough to handle other ideal metrics of non-convolution type; for example, the ideal metric \( \chi_r \) describes the rate of convergence in (1.3) with respect to the uniform metric \( \chi \); also \( \mu_r \) and \( \nu_r \), when taken together, describe the convergence rate in terms of the uniform metric \( \ell \). There are few published results concerning convergence rates in terms of \( \ell \) and \( \chi \); as for the latter metric, Banys (1976) has found convergence rates which do not hold uniformly over \( \mathbb{R} \), but only over intervals increasing with the \( n \), the sample size.

We note that Zolotarev (1976,1977) and Senatov (1980) use the metric \( \varsigma_r \) to develop the convergence rate in (1.3) with \( \alpha = 2 \) in terms of \( \text{Var} \) and \( \rho \). Even in this special case our results are sharper and more refined. Zolotarev and Rachev (1984) and Omey and Rachev (1987) use the ideal weighted Kolmogorov metric

\[
\rho_r(X_1, X_2) := \sup_{x \in \mathbb{R}} |x|^r |F_{X_1}(x) - F_{X_2}(x)|, \quad r > 1
\]

to obtain the rate of convergence for the normalized maxima

\[
n^{-1} \max(X_1, \ldots, X_n) \xrightarrow{D} U,
\]

where \( P(U < x) = \exp(-x^{-1}) \), \( x \geq 0 \).
The remainder of this section describes the special properties of the ideal convolution (or smoothing) metrics $\mu_{\theta, r}$ and $\nu_{\theta, r}$. We first verify ideality.

**Lemma 2.1.** For all $\theta \in \mathcal{X}$ and $r > 0$, $\mu_{\theta, r}$ is an ideal metric of order $r-1$.

**Proof.** If $Z$ does not depend upon $X_1$ and $X_2$ then

$$\ell(X_1 + Z, X_2 + Z) \leq \ell(X_1, X_2),$$

and hence $\mu_{\theta, r}(X_1 + Z, X_2 + Z) \leq \mu_{\theta, r}(X_1, X_2)$. Additionally, for any $c \neq 0$

$$\mu_{\theta, r}(cX_1, cX_2) = \sup_{h \in \mathbb{R}} |h|^r \ell(cX_1 + h\theta, cX_2 + h\theta)$$

$$= \sup_{h \in \mathbb{R}} |ch|^r \ell(cX_1 + ch\theta, cX_2 + ch\theta)$$

$$= |c|^{r-1} \mu_{\theta, r}(X_1, X_2).$$

Q.E.D.

The proof of the next lemma is analogous to the one above.

**Lemma 2.2.** For all $\theta \in \mathcal{X}$ and $r > 0$, $\nu_{\theta, r}$ is an ideal metric of order $r$.

We now turn to special property (P1) and show that both $\mu_{\theta, r}$ and $\nu_{\theta, r}$ are bounded above by the difference pseudomoment whenever $\theta$ has a density which is smooth enough.

**Lemma 2.3.** Let $k \in \mathbb{N}^+$ and suppose that $X, Y \in \mathcal{X}(\mathbb{R})$ satisfy $EX_j = EY_j$, $j = 1, \ldots, k-2$. Then for every $\theta \in \mathcal{X}(\mathbb{R})$ with a density $g$ which
is \( k-1 \) times differentiable,

\[
\mu_{\theta, k}(X_1, X_2) \leq \frac{||g^{(k-1)}||_{\infty}}{(k-1)!} \kappa_{k-1}(X_1, X_2),
\]

where \( || \cdot ||_{\infty} := \text{ess sup}_{x \in \mathbb{R}} (\ast) \).

**Proof.** In view of the inequality [Zolotarev (1979)]

\[
\zeta_{k-1}(X_1, X_2) \leq \frac{1}{(k-1)!} \kappa_{k-1}(X_1, X_2),
\]

it suffices to show that

\[
\mu_{\theta, k}(X_1, X_2) \leq \zeta_{k-1}(X_1, X_2).
\]

But

\[
\mu_{\theta, k}(X_1, X_2) = \sup_{h \in \mathbb{R}} |h|^k \sup_{x \in \mathbb{R}} \frac{1}{|h|} \left| \int_{-\infty}^{\infty} g(x-y) d(F_{X_1}(y) - F_{X_2}(y)) \right|
\]

\[
= \sup_{h \in \mathbb{R}} |h|^{k-1} \sup_{x \in \mathbb{R}} \left| \int_{-\infty}^{\infty} g^{(1)}(x-y) d(F_{X_1}(y) - F_{X_2}(y)) \right|
\]

\[
= \sup_{h \in \mathbb{R}} |h|^{k-2} \sup_{x \in \mathbb{R}} \left| \int_{-\infty}^{\infty} g^{(2)}(x-y) d(F_{X_1}(y) - F_{X_2}(y)) \right|
\]

\[
= \sup_{h \in \mathbb{R}} |h|^{k-3} \sup_{x \in \mathbb{R}} \left| \int_{-\infty}^{\infty} g^{(3)}(x-y) d(F_{X_1}(y) - F_{X_2}(y)) \right|
\]

\[
= \sup_{h \in \mathbb{R}} |h|^{k-4} \sup_{x \in \mathbb{R}} \left| \int_{-\infty}^{\infty} g^{(4)}(x-y) d(F_{X_1}(y) - F_{X_2}(y)) \right|
\]

\[
= \cdots
\]

\[
= \sup_{h \in \mathbb{R}} |h|^{k-k} \sup_{x \in \mathbb{R}} \left| \int_{-\infty}^{\infty} g^{(k)}(x-y) d(F_{X_1}(y) - F_{X_2}(y)) \right|
\]
where

\[(2.6) \quad F_X^{(-k+1)}(x) := \int_{-\infty}^{x} \frac{(x-t)^{k-1}}{(k-1)!} \, dF_X(t). \]

Therefore, by (1.10) and \( \zeta_{k-1} = \zeta_{k-2,1} \), we have

\[
\mu_{\theta,k}(X_1, X_2) \leq \|\sigma^{(k-1)}\|_\infty \int_{-\infty}^{\infty} |F_X^{(2-k)}(y) - F_X^{(2-k)}(y)| \, dy
\]

\[= \|\sigma^{(k-1)}\|_\infty \zeta_{k-1}(X_1, X_2). \]

Q.E.D.

Analogously, Lemmas 2.5 and 2.6 below show that \( \nu_{\theta,k} \) is bounded by the difference pseudomoments.

Under similar hypotheses we may also show that the smoothing metrics \( \mu_{\theta,k} \) and \( \nu_{\theta,k} \) are weaker than \( \zeta_{k,p} \). This, together with inequality (2.5), helps illustrate property (P2).

**Lemma 2.4.** For every \( \theta \in \mathcal{H}(\mathbb{R}) \) with a density \( g \) which is \( m \) times differentiable and for all \( X_1, X_2 \in \mathcal{H}(\mathbb{R}) \)

\[
\mu_{\theta,r}(X_1, X_2) \leq C(m, p, g) \zeta_{m-1, p}(X_1, X_2),
\]

where \( r = m + \frac{1}{p} \), \( m \in \mathbb{N}^+ \), and

\[(2.7) \quad C(m, p, g) := \|g^{(m)}\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1. \]

**Lemma 2.5.** Under the hypotheses of Lemma 2.4 we have

\[(2.8) \quad \nu_{\theta,r}(X_1, X_2) \leq C(m, p, g) \zeta_{m-1, p}(X_1, X_2), \]

\]
where \( r = m - 1 + \frac{1}{p} \), \( m \in \mathbb{N}^+ \) and \( C(m, p, g) \) is as in (2.7).

Lemma 2.6. [cf. Theorem 2 of Maejima and Rachev (1987)]. Let \( m \in \mathbb{N}^+ \) and suppose \( E(X_1^j, X_2^j) = 0 \) for \( j = 0, 1, \ldots, m-1 \). Then for \( p \in [1, \infty) \)

\[
(2.9) \quad \zeta_{m, p}(X_1, X_2) \leq \begin{cases} 
\kappa_1^{1/p}(X_1, X_2) & m = 0, \\
\frac{\Gamma(1 + \frac{1}{p})}{\Gamma(r)} \kappa_r(X_1, X_2) & r = m + \frac{1}{p}, \ m = 1, 2, \ldots
\end{cases}
\]

The proofs of the above three lemmas follow from straightforward modifications of the techniques used in Ignatov and Rachev (1983), Maejima and Rachev (1987) and Zolotarev (1979). The details are left to the reader.

§3. Rates of convergence in the total variation metric

In this section we develop rates of convergence with respect to the Var metric.

Throughout we suppose that \( X, X_1, X_2, \ldots \) denotes a sequence of i.i.d. random variables in \( \mathcal{X}(B) \), where \( B \) is a separable Banach space. \( Y \in \mathcal{X}(B) \) denotes a strictly \( \alpha \)-stable random variable. The ideal convolution metric \( \nu_r := \nu_{\alpha, r} \) (i.e., \( \theta = Y \)) will play a central role.

Our main theorem is

**Theorem 3.1.** Let \( Y \) be an \( \alpha \)-stable random variable. Let

\[ r = s + \frac{1}{p} > \alpha \]

for some integer \( s \) and \( p \in (1, \infty) \), \( a = \frac{1}{2^{r/\alpha} A} \),

and \( A := 2^{2^r/\alpha - 1} + 3^{r/\alpha} \). If \( X \in \mathcal{X}(B) \) satisfies
\[ (3.1) \quad \tau_0 := \tau_0(X,Y) := \max(\operatorname{Var}(X,Y), \nu_{\alpha,r}(X,Y)) \leq a, \]

then \( \forall n \geq 1 \)

\[ \operatorname{Var} \left( \frac{X_1 + \ldots + X_n}{n^{1/\alpha}}, Y \right) \leq A(a) \tau_0 n^{1-r/\alpha} \leq 2^{-r/\alpha} n^{1-r/\alpha}. \]

Remarks

(i) A result of this type was proved by Senatov (1980) for the special case \( B = \mathbb{R}^k \), \( s = 3 \), and \( \alpha = 2 \) via the \( \varsigma_r \) metric. We will follow Senatov's method with some refinements.

(ii) Theorem 3.1 is optimal in the sense that the power of \( n \) is the smallest possible; i.e., the exponent \( 1-r/\alpha \) cannot be decreased. This follows, for example, from Theorem 3.4.1 of Ibragimov and Linnik (1971) and the inequality \( \rho \leq \sigma = \frac{1}{2} \operatorname{Var} \in \mathcal{X}(\mathbb{R}) \).

Before proving Theorem 3.1 we need a few auxiliary results.

Lemma 3.2. For any \( X_1, X_2 \in \mathcal{X}(B) \) and \( \sigma > 0 \)

\[ \operatorname{Var}(X_1 + \sigma Y, X_2 + \sigma Y) \leq \sigma^{-r} \nu_r(X_1, X_2). \]

Proof. Since \( Y \) and \((-Y)\) have the same distribution

\[ \nu_r(X_1, X_2) = \sup_{h>0} h^r \ell_1(X_1 + hY, X_2 + hY) \]

and thus

\[ \ell_1(X_1 + hY, X_2 + hY) \leq h^{-r} \sup_{h>0} h^r \ell_1(X_1 + hY, X_2 + hY) \]
Lemma 3.2 closely resembles Lemma 1 of Senatov (1980) for the metric $s_r$. The next result resembles Lemma 2 of Senatov (1980) proved for $B = \mathbb{R}^k$. We note that estimates of this sort have been used by Sazonov (1972) and Sazonov and Ul'yanov (1979).

**Lemma 3.3.** For any $X_1, X_2, U, V \in \mathcal{X}(B)$ the following inequality holds:

$$\text{Var}(X_1 + U, X_2 + U) \leq \text{Var}(X_1, X_2)\text{Var}(U, V) + \text{Var}(X_1 + V, X_2 + V).$$

**Proof.** By the definition (1.5) and the triangle inequality

$$\text{Var}(X_1 + U, X_2 + U) = \sup\{|Ef(X_1 + U) - Ef(X_2 + U)| : ||f||_\infty \leq 1\}$$

$$= \sup\{|\int_B f(u)(P_{X_1+U}-P_{X_2+U}) \, du| : ||f||_\infty \leq 1\}$$

$$\leq \sup\{|\int_B \tilde{f}(x)(P_{X_1}-P_{X_2}) \, dx| : ||f||_\infty \leq 1\} +$$

$$+ \text{Var}(X_1 + U, X_2 + V),$$

where

$$\tilde{f}(x) := \int_B f(u)(P_U-P_V)(du-x) = \int_B f(u+x)(P_U-P_V) \, du,$$

and where $P_X$ denotes the law of the random variable $X$. Since $||f||_\infty \leq 1$ then

$$||\tilde{f}||_\infty = \sup_{x \in B} \left| \int_B f(u+x)(P_U-P_V) \, du \right|$$

$$\leq \left| \int_B f(u)(P_U-P_V) \, du \right|$$
\[ \leq \text{Var}(U,V), \text{ by (1.5)}, \]

and thus \( \sup \left\{ \left| \int_B f(x)(P_{X_1} - P_{X_2}) dx \right| : \|f\|_\infty \leq 1 \right\} \) is bounded by

\[ \leq \sup \left\{ \left| \int_B g(x)(P_{X_1} - P_{X_2}) dx \right| : \|g\|_\infty \leq \text{Var}(U,V) \right\} \]

\[ = \text{Var}(X_1, X_2) \text{Var}(U,V). \]

Q.E.D.

We now prove Theorem 3.1; throughout, \( Y, Y_1, Y_2, \ldots \) denote i.i.d. copies of \( Y \).

**Proof.** We proceed by induction; for \( n = 1 \) the assertion of the theorem is trivial. For \( n = 2 \), the assertion follows from the inequality

\[ \text{Var}(\frac{X_1 + X_2}{2^{1/\alpha}}, Y) = \text{Var}(\frac{X_1 + X_2}{2^{1/\alpha}}, \frac{Y_1 + Y_2}{2^{1/\alpha}}) \]

\[ = \text{Var}(X_1 + X_2, Y_1 + Y_2) \]

\[ \leq 2 \text{Var}(X_1, Y_1) \]

\[ \leq A(a) r_0 2^{1-\tau/\alpha} \]

since \( A(a) \geq 2^{\tau/\alpha} \). A similar calculation holds for \( n = 3 \). Suppose now that the estimate

\[ (3.2) \quad \text{Var}(\frac{X_1 + \ldots + X_j}{j^{1/\alpha}}, Y) \leq A(a) r_0^{1-\tau/\alpha} \]

holds for all \( j < n \). To complete the induction we only need to show that \( (3.2) \) holds for \( j = n \).

Thus, assuming \( (3.2) \) we have by \( (3.1) \)
For any integer \( n \geq 4 \) and \( m = \lfloor \frac{n}{2} \rfloor \), where \([\cdot]\) denotes integer part, the triangle inequality gives

\[
V := \text{Var}(\frac{X_1 + \ldots + X_n}{n^{1/\alpha}}, Y)
\]

\[
= \text{Var}(\frac{X_1 + \ldots + X_n}{n^{1/\alpha}}, \frac{Y_1 + \ldots + Y_n}{n^{1/\alpha}})
\]

\[
\leq \text{Var}(\frac{X_1 + \ldots + X_m}{n^{1/\alpha}}, \frac{X_{m+1} + \ldots + X_n}{n^{1/\alpha}}, \frac{Y_1 + \ldots + Y_m}{n^{1/\alpha}}, \frac{Y_{m+1} + \ldots + Y_n}{n^{1/\alpha}})
\]

\[
+ \text{Var}(\frac{X_1 + \ldots + X_m}{n^{1/\alpha}}, \frac{X_{m+1} + \ldots + X_n}{n^{1/\alpha}}, \frac{Y_1 + \ldots + Y_m}{n^{1/\alpha}}, \frac{Y_{m+1} + \ldots + Y_n}{n^{1/\alpha}})
\]

Hence, by Lemma 3.3

\[
(3.4) \quad V \leq I_1 + I_2 + I_3,
\]

where

\[
I_1 := \text{Var}(\frac{X_1 + \ldots + X_m}{n^{1/\alpha}}, \frac{Y_1 + \ldots + Y_m}{n^{1/\alpha}}) + \text{Var}(\frac{X_{m+1} + \ldots + X_n}{n^{1/\alpha}}, \frac{Y_{m+1} + \ldots + Y_n}{n^{1/\alpha}})
\]

\[
I_2 := \text{Var}(\frac{X_1 + \ldots + X_m}{n^{1/\alpha}} + \frac{X_{m+1} + \ldots + X_n}{n^{1/\alpha}}, \frac{Y_1 + \ldots + Y_m}{n^{1/\alpha}} + \frac{Y_{m+1} + \ldots + Y_n}{n^{1/\alpha}})
\]
\[ I_3 := \text{Var}\left( \frac{Y_1 + \ldots + Y_m}{n^{1/\alpha}} + \frac{X_{m+1} + \ldots + X_n}{n^{1/\alpha}}, \frac{Y_1 + \ldots + Y_m}{n^{1/\alpha}} + \frac{Y_{m+1} + \ldots + Y_n}{n^{1/\alpha}} \right). \]

We first estimate \( I_1 \). By (3.3) and (3.2)

\[
I_1 \leq 2^{-r/\alpha} A(a)r_0(n-m)^{1-r/\alpha}
\]

(3.5)

\[
\leq \frac{1}{2} A(a)r_0 n^{1-r/\alpha}.
\]

In order to estimate \( I_2 \) and \( I_3 \) we will use Lemma 3.2 and the relation

\[
\frac{Y_1 + \ldots + Y_n}{n^{1/\alpha}} \equiv Y_1.
\]

(3.6)

Thus, by (3.6), Lemma 3.2, and the fact that \( \nu_r \) is ideal of order \( r \) we deduce

\[
I_2 = \text{Var}\left( \frac{X_1 + \ldots + X_m}{n^{1/\alpha}} + \left( \frac{n-m}{n} \right)^{1/\alpha} \frac{Y_1 + \ldots + Y_m}{n^{1/\alpha}} \right)
\]

\[
\leq \left( \frac{n-m}{n} \right)^{-r/\alpha} \nu_r\left( \frac{X_1 + \ldots + X_m}{n^{1/\alpha}}, \frac{Y_1 + \ldots + Y_m}{n^{1/\alpha}} \right)
\]

\[
\leq 2^{r/\alpha} \nu_r\left( \frac{X_1}{n^{1/\alpha}}, \frac{Y_1}{n^{1/\alpha}} \right)
\]

(3.7)

\[
\leq 2^{r/\alpha-1} n^{1-r/\alpha} \nu_r(X_1, Y_1).
\]
Analogously, we estimate $I_3$ by

$$I_3 = \text{Var}(\frac{X_1 + \ldots + X_{n-m}}{n^{1/\alpha}} + \frac{(m)}{n} Y, \frac{Y_1 + \ldots + Y_{n-m}}{n^{1/\alpha}} + \frac{(m)}{n} Y)$$

$$\leq \frac{m}{n} - \frac{r}{\alpha} \nu_r (\frac{X_1 + \ldots + X_{n-m}}{n^{1/\alpha}}, \frac{Y_1 + \ldots + Y_{n-m}}{n^{1/\alpha}})$$

(3.8) \hspace{1cm} \leq 3^{r/\alpha} n^{1-r/\alpha} \nu_r (X_1, Y_1).

Taking (3.4), (3.5), (3.7) and (3.8) into account we obtain

$$V \leq \left(\frac{1}{2} A(a) + 2^{r/\alpha-1} + 3^{r/\alpha}\right) r_0 n^{1-r/\alpha}$$

$$\leq A(a) r_0 n^{1-r/\alpha},$$

since $A(a)/2 = 2^{r/\alpha-1} + 3^{r/\alpha}.$

Q.E.D.

§4. Rates of convergence in the $\chi$ metric

In this section we develop rates of convergence in (1.3) with respect to the $\chi$ metric; our purpose here is to show that the methods of proof for Theorem 3.1 can be easily extended to deduce analogous results with respect to $\chi$. The metric $\chi_r$ will play a role analogous to that played by $\nu_r$ in Section 3.

Throughout, $X, X_1, X_2, \ldots$ denotes a sequence of i.i.d. copies of an $\alpha$-stable random variable.

Our main theorem is

**Theorem 4.1.** Let $Y$ be an $\alpha$-stable random variable in $\mathcal{X}(\mathbb{R})$. 

Let \( r > \alpha, \quad b := \frac{1}{2^{r/\alpha}B} \), and \( B := \max(3^r/\alpha, 2C_r(2^r/\alpha - 1 + 3^r/\alpha)) \) where
\[
C_r := \frac{r}{\frac{r}{\alpha}e}.
\]
If \( X \in \mathcal{X}(\mathbb{R}) \) satisfies
\[
\tau_r := \tau_r(X,Y) := \max(\chi(X,Y), \chi_r(X,Y)) \leq b,
\]
then for all \( n \geq 1 \)
\[
\chi_{(-\frac{1}{n^{1/\alpha}}, \frac{1}{n^{1/\alpha}}), Y} \leq Br n^{1-r/\alpha} = 2^{-r/\alpha} n^{1-r/\alpha}.
\]

Remarks

(i) In comparing conditions (3.1) and (4.1) it is useful to note that the metric \( \chi \) is topologically weaker than \( \text{Var} \), i.e., \( \text{Var}(X_n, Y) \to 0 \) implies \( \chi(X_n, Y) \to 0 \) but not conversely. Also, it is easy to show that if \( r = m + \beta, \quad m = 0,1,\ldots, \beta \in (0,1) \)
then
\[
\chi_r \leq C_\beta \chi_r \quad \text{where} \quad C_\beta := \sup \{|t\beta(1-e^{-it})|:}
\]
if \( r = m, \quad m = 1,2,\ldots \) then \( \chi_r \leq \chi_r \).

(ii) Actually, one may show that for \( r \in \mathbb{N}^+ \) the metric \( \chi_r \) has a convolution type structure. In fact, with a slight abuse of notation,
\[
\chi_r(F_{X_1}, F_{X_2}) = \chi(F_{X_1}^* p_r, F_{X_2}^* p_r).
\]

\[p_r(t) = t^{r/r!} 1_{\{t > 0\}} \] is the density of an unbounded positive measure on the half line \([0, \infty)\).

(iii) As in Theorem 3.1, the exponent \( 1-r/\alpha \) cannot be reduced; this follows from Theorem 3.4.1 of Ibragimov and Linnik (1971) and
the fact that \( \chi \) convergence implies \( \rho \) convergence.

(iv) As noted earlier, Banys (1976) has obtained a result similar to Theorem 4.1: his result is weaker since it only considers the sup norm difference between characteristic functions over finite intervals depending on \( n \). Additionally, his result is expressed in terms of the so-called \( r \)th absolute pseudomoment

\[
\nu_r(X,Y) := \int |p_X(x) - p_Y(x)| \, dx;
\]

since \( \chi_r \) is topologically weaker than \( \nu_r \) (i.e., \( \chi_r \leq \nabla_r \)), \( \zeta_r \leq \nabla_r \), and \( \chi_r \) convergence does not, in general, imply \( \nu_r \) convergence, our estimate (4.2) is clearly more refined, even over finite intervals.

The proof of Theorem 4.1 is very similar to that of Theorem 3.1 and uses the following auxiliary results, which are completely analogous to Lemmas 3.2 and 3.3. We leave the details to the reader.

**Lemma 4.2.** For any \( X_1, X_2 \in \mathcal{X}(\mathbb{R}), \sigma > 0, \) and \( r > \alpha \)

\[
\chi(X_1 + \sigma Y, X_2 + \sigma Y) \leq C_r \sigma^r \chi_r(X_1, X_2),
\]

where \( C_r := \left( \frac{r}{\alpha e} \right)^{r/\alpha} \).

**Proof.** We have

\[
\chi(X_1 + \sigma Y, X_2 + \sigma Y) := \sup_{t \in \mathbb{R}} |\phi_{X_1}(t) - \phi_{X_2}(t)| \phi_{\sigma Y}(t)
\]
\begin{align*}
&= \sup_{t \in \mathbb{R}} |\varphi_{\chi_1} (t) - \varphi_{\chi_2} (t)| \exp \{-|\sigma t|^\alpha\} \\
&\leq \sup_{t \in \mathbb{R}} |\sigma t|^{-r} |\varphi_{\chi_1} (t) - \varphi_{\chi_2} (t)| \sup_{u > 0} u^{-u^\alpha} \\
&= C_r \sigma^{-r} \chi_r (X, Y),
\end{align*}

since $C_r = \sup_{u > 0} u^{-u^\alpha}$ by a simple computation.

Q.E.D.

Lemma 4.3. For any $X_1, X_2, Z, W \in \mathcal{H}(\mathbb{R})$ the following inequality holds:

$$
\chi(X_1 + Z, X_2 + W) \leq \chi(X_1, X_2) \chi(Z, W) + \chi(X_1 + W, X_2 + W).
$$

Proof. From the inequality

$$
|\varphi_{X_1 + Z}(t) - \varphi_{X_2 + W}(t)| \leq |\varphi_{X_1}(t) - \varphi_{X_2}(t)| + |\varphi_{Z}(t) - \varphi_{W}(t)| + \\
+ |\varphi_{X_1}(t) - \varphi_{X_2}(t)| |\varphi_{W}(t)|
$$

we obtain the desired result.

Q.E.D.

§5. Rates of convergence in the $\ell$ metric

In this section we develop convergence rates with respect to the $\ell$-metric and thus we naturally restrict attention to the subset $\mathcal{H}^*$ of $\mathcal{H}(\mathbb{R}^k)$ of random variables with densities. Throughout $X, X_1, X_2, \ldots$ denotes a sequence of i.i.d. random variables in $\mathcal{H}^*$ and $Y = Y_\alpha$
denotes an $\alpha$-stable random variable. The ideal convolution metrics
$\mu_r := \mu_{\alpha, r}$ and $\nu_r := \nu_{\alpha, r}$ (i.e., $\theta = Y$) will play a central role.

Our main result is

**Theorem 5.1.** Let $Y$ be an $\alpha$-stable random variable in $\mathcal{X}(\mathbb{R}^k)$. Let

$$r = m + \frac{1}{p} > \alpha$$

for some integer $m$ and $p \in [1, \infty)$, $a := \frac{1}{2^{r/\alpha} A}$,

$$A := 2^{(r-\alpha) - 3(r+1)/\alpha}$$

and $D := 3^{1/\alpha} 2^{r/\alpha}$. If $X \in \mathbb{X}^*$ satisfies

(i) $\tau(X, Y) := \max(\ell(X, Y), \mu_{\alpha, r}(X, Y)) \leq a$ and

(ii) $\tau_0(X, Y) := \max(\text{Var}(X, Y), \nu_{\alpha, r}(X, Y)) \leq \frac{1}{A(a) D} \leq a$,

then

$$(5.1)$$

$$\ell\left(\frac{X_1 + \ldots + X_n}{n^{1/\alpha}}, Y\right) \leq A(a) \tau(X, Y)n^{1-r/\alpha}.$$ 

**Remarks.**

(i) Conditions (i) and (ii) describe the domain of attraction of a stable $Y_\alpha$ random variable; in fact, they guarantee $\ell$-closeness (of order $n^{1-r/\alpha}$) between $Y$ and the normalized sums

$$(5.3)$$

$$n^{-1/\alpha}(X_1 + \ldots + X_n).$$

(ii) From property (P2) and especially Lemmas 2.3, 2.5 and 2.6 we know that $\mu_{r+1}(X, Y)$ and $\nu_r(X, Y)$, $r = m - 1 + \frac{1}{p}$, $m = 1, 2, \ldots$ can be approximated from above by the $h$th difference pseudomoment $\kappa_r$ whenever $X$ and $Y$ share the same first $(m-1)$ moments. Thus conditions (i)
and (ii) could be expressed in terms of difference pseudomoments, which of course amounts to conditions on the tails of $X$.

(iii) That (5.2) is of the right order of magnitude may be seen from Theorem 4.5.1 of Ibragimov and Linnik (1971).

To prove Theorem 5.1 we need a few auxiliary results similar in spirit to Lemmas 3.2 and 3.3.

**Lemma 5.2.** Let $X_1, X_2 \in \mathcal{X}(\mathbb{R}^k)$. Then

$$\ell(X_1+\sigma Y, X_2+\sigma Y) \leq \sigma^{-r} \mu_r(X_1, X_2).$$

**Proof:**

$$\ell(X_1+\sigma Y, X_2+\sigma Y) \leq \sigma^{-r} \ell(X_1+\sigma Y, X_2+\sigma Y) \leq \sigma^{-r} \mu_r(X_1, X_2).$$

Q.E.D.

**Lemma 5.3.** For any $X,Y,U,V \in \mathcal{X}(\mathbb{R}^k)$ the following inequality holds:

$$\ell(X+U,Y+U) \leq \ell(X,Y)\text{Var}(U,V) + \ell(X+V,Y+V)$$

**Proof:** Using the triangle inequality we obtain

$$\ell(X+U,Y+U) = \sup_{x \in \mathbb{R}} |\int (p_X(x-y) - p_Y(x,y)) \Pr(U \in dy) |$$

$$\leq \sup_{x \in \mathbb{R}} |\int (p_X(x,y) - p_Y(x,y)) (\Pr(U \in dy) - \Pr(V \in dy)) | +$$

$$+ \sup_{x \in \mathbb{R}} |\int (p_X(x-y) - p_Y(x-y)) \Pr(V \in dy) |$$

$$\leq \ell(X,Y)\text{Var}(U,V) + \ell(X+V,Y+V).$$
To prove Theorem 5.1 one only needs to use the method of proof for Theorem 3.1 combined with the above two auxiliary results. The complete details are left to the reader.

§6. The ideal smoothing metrics and weak convergence

We conclude our discussion of the ideal metrics $\mu_r$ and $\nu_r$ by showing that they satisfy the same weak convergence properties as do the Kantorovich-Wasserstein distance $W_\alpha$ and the pseudomoments $\kappa_r$.

**Theorem 6.1** Let $k \in \mathbb{N}^+$, $0 < \alpha \leq 2$, and $X_n, U \in \mathcal{E}(\mathbb{R})$ with $EX_n^j = EU^j \forall j = 1, \ldots, k-2$. If $k$ is odd then the following are equivalent as $n \to \infty$:

1. $\mu_{\alpha,k}(X_n, U) \to 0$,
2. (a) $X_n \Rightarrow U$ and (b) $E|X_n|^{k-1} \to E|U|^{k-1}$,
3. $W_{k-1}(X_n, U) \to 0$,
4. $\kappa_{k-1}(X_n, U) \to 0$, and
5. $\nu_{\alpha,k-1}(X_n, U) \to 0$.

Before proving this we note that (ii) $\iff$ (iii) follows immediately from Theorem 4.1 of Rachev (1984 c), Theorem 1 of Rachev (1984 b), Theorem 2 of Rachev (1984 a) and the identity $EX^{k-1} = E|X|^{k-1}$ for $k$ odd. Also, (ii) $\iff$ (iv) follows from Rachev (1982); (iv) $\Rightarrow$ (i) follows from Lemma 2.3, and (iv) $\Rightarrow$ (v) from Lemmas 2.5 and 2.6; thus the only new result here are the implications (i) $\Rightarrow$ (ii) and (v) $\Rightarrow$ (ii).

Now (i) $\Rightarrow$ (ii)(a) follows easily from Fourier transform arguments since the Fourier transform of $g$ never vanishes. Similarly if (v)
holds then \( X_n + Y \to U + Y \) and thus (ii)(a) follows. To prove

(i) \(\Rightarrow\) (ii)(b) we need a lemma.

**Lemma 6.2.** Let \( 0 < \alpha \leq 2 \) and consider the associated metric

\[ \mu_r := \mu_{r, \alpha}. \]

For all \( k \) there is a constant \( \beta := \beta(\alpha, k) < \infty \) such that for all \( X, U \in \mathcal{X}(\mathbb{R}) \)

\[ (6.1) \quad \mu_k(X, U) \geq \beta \left| \int_{-\infty}^{\infty} F_X^{(2-k)}(z) - F_U^{(2-k)}(z) \, dz \right|. \]

Here \( F^{(2-k)} \) is as in (2.6).

**Proof of Theorem 6.1.** Using equality of the first \( k-2 \) moments and applying (6.1) to \( X_n \) and \( U \) yields

\[ \beta^{-1} \mu_k(X_n, U) \geq \left| \int_{-\infty}^{\infty} \int_{-\infty}^{z} \frac{(z-t)^{k-2}}{(k-2)!} (dF_{X_n} - dF_U)(t) \, dt \, dz \right| \]

\[ = \left| \int_{-\infty}^{0} (\ast) \, dt + \int_{0}^{\infty} (\ast) \, dt \right| \]

\[ (6.2) \quad := \left| I_1 + I_2 \right|. \]

To estimate \( I_1 \) and \( I_2 \) we first note that since

\[ \int_{-\infty}^{\infty} \frac{(z-t)^{k-2}}{(k-2)!} (dF_{X_n} - dF_U)(t) = \mathbb{E}(z-X_n)^{k-2} - \mathbb{E}(z-U)^{k-2} = 0, \]

we obtain
Thus by (6.3) and Fubini's theorem we get

\[
I_2 = (-1)^{k-1} \int_0^\infty \int_0^t \frac{(t-z)^{k-2}}{(k-2)!} (dF_X - dF_U)(t) \, dz

= (-1)^{k-1} \int_0^t \frac{(t-z)^{k-2}}{(k-2)!} \, dz \, d(F_X - F_U)(t)

(6.4)
\[
= \int_0^\infty \frac{(-t)^{k-1}}{(k-1)!} (dF_X - dF_U)(t) .
\]

Another application of Fubini's theorem gives

\[
I_1 = \int_{-\infty}^0 \int_t^\infty \frac{(z-t)^{k-2}}{(k-2)!} \, dz \, (dF_X - dF_U)(t)

= \int_{-\infty}^0 \frac{(-t)^{k-1}}{(k-1)!} \, (dF_X - dF_U)(t) .

(6.5)
\]

Combining (6.3), (6.4) and (6.5) gives

\[
\beta^{-1} \mu_k(X_n, U) \geq \left| \int_{-\infty}^\infty \frac{(-t)^{k-1}}{(k-1)!} (dF_X - dF_U)(t) \right|
\]
which gives the desired implication (i) \Rightarrow (ii)(b).

To prove (v) \Rightarrow (ii)(b) we integrate by parts to obtain

\[ \nu_k(X_n, U) \geq \int_{-\infty}^{\infty} \left| p_{X_n+Y}(x) - p_{U+Y}(x) \right| dx \]

\[ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_Y(z-x) \left( \int_{z}^{\infty} \frac{(z-t)^{k-1}}{(k-1)!} d(F_{X_n}(t) - F_{U}(t)) \right) dx \]

\[ \geq \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_Y(z-x) dx \int_{-\infty}^{z} \frac{(z-t)^{k-1}}{(k-1)!} d(F_{X_n}(t) - F_{U}(t)) dz \right| \]

\[ = \left| \int_{-\infty}^{\infty} \int_{-\infty}^{z} \frac{(z-t)^{k-1}}{(k-1)!} d(F_{X_n}(t) - F_{U}(t)) dz \right| \cdot \left| \int_{-\infty}^{\infty} p_Y(x) dx \right| \]

By (6.2) – (6.5) we obtain

\[ \nu_k(X_n, U) \geq \left| \int_{-\infty}^{\infty} p_Y(x) dx \right| \cdot \left| E(\frac{X_n^k-U^k}{k}) \right| , \]

showing (v) \Rightarrow (ii)(b) and completing Theorem 6.1.

Q.E.D.

It only remains to give the

Proof of Lemma 6.2. Integration by parts yields
\[
\mu_k(X,U) = \sup_{h \in \mathbb{R}} \left| h^k \sup_{x \in \mathbb{R}} \left| p_{x+hY}^{(x)} - p_{U+hY}^{(x)} \right| \right|
\]
\[
= \sup_{h \in \mathbb{R}} \left| h^k \sup_{x \in \mathbb{R}} \left| \int_{-\infty}^{\infty} p_{hY}^{(z)} d(F_X^{(x-z)} - F_U^{(x-z)}) \right| \right|
\]
\[
(6.6) = \sup_{h \in \mathbb{R}} \left| h^k \sup_{x \in \mathbb{R}} \left| \int_{-\infty}^{\infty} F_X^{(2-k)}(x-z) - F_U^{(2-k)}(x-z) p_{hY}^{(k-1)}(z) dz \right| \right|
\]

Now \( 2\pi p_{hY}^{(z)} = \int_{-\infty}^{\infty} e^{-itz} e^{-|ht|^\alpha} dt \) and differentiating \( p_{hY}^{(k-1)} \) times gives (setting \( \bar{t} = th \)):

\[
2\pi \left| h^k p_{hY}^{(k-1)}(z) \right| = \left| h^k \int_{-\infty}^{\infty} (it)^{k-1} e^{itz} |ht|^\alpha dt \right|
\]
\[
= \left| h^k \int_{-\infty}^{\infty} (i\bar{t})^{k-1} e^{i\bar{t}z/h} |\bar{t}|^\alpha d(\frac{\bar{t}}{h}) \right|
\]
\[
= \left| \int_{-\infty}^{\infty} (it)^{k-1} e^{itz/h} |t|^\alpha dt \right|
\]

Since \( \beta := \beta(\alpha, k) := \frac{1}{2\pi} \int_{-\infty}^{\infty} |it|^{k-1} e^{-|t|^\alpha} dt < \infty \) we obtain
\[
\lim_{h \to \infty} \left| h^{-p(k-1)}(z) \right| = \frac{1}{2\pi} \left| \int_{-\infty}^{\infty} \lim_{h \to \infty} \left( \frac{|it|^{k-1} e^{izh}}{|h-|t|} \right) dt \right| = \beta.
\]

Now multiply both sides of (6.6) by \( \beta^{-1} \). Since

\[
\int_{-\infty}^{\infty} |t|^{k-1} e^{-|t|^\alpha} dt \text{ and } \int |F_X(x-z)-F_U(x-z)| dt = \gamma_{k-1}(X,U)
\]

both finite, \( \beta^{-1} \mu_k(X,U) \) is

\[
\geq \beta^{-1} \sup_{x \in \mathbb{R}} \left| \int_{-\infty}^{\infty} F_X^{(2-k)}(x-z)-F_U^{(2-k)}(x-z) \lim_{h \to \infty} \left( h^{-p(k-1)}(z) \right) dz \right|
\]

\[
= \sup_{x \in \mathbb{R}} \left| \int_{-\infty}^{\infty} F_X^{(2-k)}(x-z)-F_U^{(2-k)}(x-z) \right| dz
\]

\[
= \sup_{x \in \mathbb{R}} \left| \int_{-\infty}^{\infty} F_X^{(2-k)}(z)-F_U^{(2-k)}(z) \right| dz.
\]

Q.E.D.

§7. Concluding remarks

The results above show that the "ideal" structure of the
convolution metrics \( \mu_r \) and \( \nu_r \) may be used to determine optimal
rates of convergence in the general central limit theorem problem. The
rates are expressed in terms of the uniform metrics \( \text{Var}, \chi \) and \( \ell \)
and hold uniformly in \( n \) under the sufficient conditions (3.1), (4.1),
and (5.1), respectively. We have not explored the possible weakening
of these conditions or even their possible necessity. This would be an
interesting line of future research.

The ideal convolution metrics \( \mu_r \) and \( \nu_r \) are not limited to the
context of Theorems 3.1, 4.1 and 5.1, but they can also be successfully employed to study other questions of interest. For example, we only mention here that $\nu_r$ can be used to prove a Berry–Esseen type of estimate for the Kolmogorov metric $\rho$ (1.8).

More precisely, if $X, X_1, X_2, \ldots$ denotes a sequence of i.i.d random variables in $\mathbb{R}$ and $Y \in \mathbb{R}$ an $\alpha$-stable random variable, then for all $r > \alpha$ and $n \geq 1$

$$
\rho\left(\frac{X_1 + \ldots + X_n}{n^{1/\alpha}}, Y\right) \leq C\nu_r(X, Y)n^{1-r/\alpha} + C\max\{\rho(X, Y), \nu_{1,1}(X, Y), \nu_{\alpha,\alpha}(X, Y)\}n^{1-\alpha/\alpha},
$$

where $C$ is an absolute constant.

Clearly, whenever $\nu_{1,1}(X, Y) < \infty$ and $\nu_{\alpha,\alpha}(X, Y) < \infty$ we obtain the right order estimate in the Berry–Esseen theorem in terms of the metric $\nu_{\alpha,\alpha}$. Inequalities of this type have been proved by Paulauskas (1982), who uses the $\rho$ metric, and by Senatov (1981), who uses $\sigma$ instead of $\rho$ and who only considers the normal case $\alpha = 2$ together with the $\zeta_{2,1}$ metric. The estimate (7.1), which proceeds by induction and which will be detailed in a forthcoming article, represents an improvement over earlier estimates since the $\nu_{\alpha,\alpha}$ distance is weaker than the $\zeta_{m,p}$ distance ($r = m-1+\frac{1}{p}$); see e.g. Lemma 2.5.

Thus metrics of the convolution type, especially those with the "ideal" structure, are extremely appropriate when investigating sums of independent random variables converging to a stable limit law. We can only conjecture that there are other ideal convolution metrics, other than those explored in this article, which may furnish additional results in related limit theorem problems.

References

Banys, J. (1976) On the convergence rate in the multidimensional local


S. T. Rachev
Department of Applied Mathematics and Statistics
State University of New York at Stony Brook
Stony Brook, New York 11794-3651

J. E. Yukich
Department of Mathematics
Lehigh University
Bethlehem, PA 18015
END
DATE
FILMED
MARCH 1988
DTIC