Bernoulli **11**(5), 2005, 829–845

Central limit theorem and convergence to stable laws in Mallows distance

OLIVER JOHNSON* and RICHARD SAMWORTH

*Statistical Laboratory, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK. E-mail: *o.johnson@statslab.cam.ac.uk*

We give a new proof of the classical central limit theorem, in the Mallows (L^r-Wasserstein) distance. Our proof is elementary in the sense that it does not require complex analysis, but rather makes use of a simple subadditive inequality related to this metric. The key is to analyse the case where equality holds. We provide some results concerning rates of convergence. We also consider convergence to stable distributions, and obtain a bound on the rate of such convergence.

Keywords: central limit theorem; Mallows distance; probability metric; stable law; Wasserstein distance

1. Introduction and main results

The spirit of the central limit theorem, that normalized sums of independent random variables converge to a normal distribution, can be understood in different senses, according to the distance used. For example, in addition to the standard central limit theorem in the sense of weak convergence, we mention the proofs in Prohorov (1952) of L^1 convergence of densities, in Gnedenko and Kolmogorov (1954) of L^∞ convergence of densities, in Barron (1986) of convergence in relative entropy and in Shimizu (1975) and Johnson and Barron (2004) of convergence in Fisher information. In this paper we consider the central limit theorem with respect to the Mallows distance and prove convergence to stable laws in the infinite-variance setting. We study the rates of convergence in both cases.

Definition 1.1. *For any $r > 0$, we define the Mallows r-distance between probability distribution functions F_X and F_Y as*

$$d_r(F_X, F_Y) = \left(\inf_{(X,Y)} \mathbb{E}|X - Y|^r \right)^{1/r},$$

where the infimum is taken over pairs (X, Y) whose marginal distribution functions are F_X and F_Y respectively, and may be infinite. Where it causes no confusion, we write $d_r(X, Y)$ for $d_r(F_X, F_Y)$.

Define $\mathcal{F}_r$ to be the set of distribution functions F such that $\int |x|^r \mathrm{d}F(x) < \infty$. Bickel and Freedman (1981) show that for $r \geqslant 1$, d_r is a metric on $\mathcal{F}_r$. If $r < 1$, then d_r^r is a metric on $\mathcal{F}_r$. In considering stable convergence, we shall also be concerned with the case where the absolute rth moments are not finite.

Throughout the paper, we write Z_{μ,σ^2} for a $N(\mu, \sigma^2)$ random variable, Z_{σ^2} for a $N(0, \sigma^2)$ random variable, and Φ_{μ,σ^2} and Φ_{σ^2} for their respective distribution functions. We establish the following main theorems:

Theorem 1.1. *Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean zero and finite variance $\sigma^2 > 0$, and let $S_n = (X_1 + \ldots + X_n)/\sqrt{n}$. Then*

$$\lim_{n\to\infty} d_2(S_n, Z_{\sigma^2}) = 0.$$

Moreover, Theorem 3.2 below shows that for any $r \geqslant 2$, if $d_r(X_i, Z_{\sigma^2}) < \infty$, then $\lim_{n\to\infty} d_r(S_n, Z_{\sigma^2}) = 0$. Theorem 1.1 implies the standard central limit theorem in the sense of weak convergence (Bickel and Freedman 1981, Lemma 8.3).

Theorem 1.2. *Fix $\alpha \in (0, 2)$, and let $X_1, X_2, \ldots$ be independent random variables (where $\mathbb{E}X_i = 0$, if $\alpha > 1$), and $S_n = (X_1 + \ldots + X_n)/n^{1/\alpha}$. If there exists an α-stable random variable Y such that $\sup_i d_\beta(X_i, Y) < \infty$ for some $\beta \in (\alpha, 2]$, then $\lim_{n\to\infty} d_\beta(S_n, Y) = 0$. In fact*

$$d_\beta(S_n, Y) \leqslant \frac{2^{1/\beta}}{n^{1/\alpha}} \left(\sum_{i=1}^{n} d_\beta^\beta(X_i, Y) \right)^{1/\beta},$$

so in the identically distributed case the rate of convergence is $O(n^{1/\beta - 1/\alpha})$.

See also Rachev and Rüschendorf (1992, 1994), who obtain similar results using different techniques in the case of identically distributed X_i and strictly symmetric Y. In Lemma 5.1 below we exhibit a large class $\mathcal{C}_K$ of distribution functions F_X for which $d_\beta(X, Y) \leqslant K$, so the theorem can be applied.

Theorem 1.1 follows by understanding the subadditivity of $d_2^2(S_n, Z_{\sigma^2})$ (see equation (4)). We consider the powers-of-two subsequence $T_k = S_{2^k}$, and use Rényi's method, introduced in Rényi (1961) to provide a proof of convergence to equilibrium of Markov chains; see also Kendall (1963). This technique was also used in Csiszár (1965) to show convergence to Haar measure for convolutions of measures on compact groups, and in Shimizu (1975) to show convergence of Fisher information in the central limit theorem. The method has four stages:

1. Consider independent and identically distributed random variables X_1 and X_2 with mean μ and variance $\sigma^2 > 0$, and write $D(X)$ for $d_2^2(X, Z_{\mu,\sigma^2})$. In Proposition 2.3, we observe that

$$D\left(\frac{X_1 + X_2}{\sqrt{2}}\right) \leqslant D(X_1), \tag{1}$$

   with equality if and only if $X_1, X_2 \sim Z_{\mu,\sigma^2}$. Hence $D(T_k)$ is decreasing and bounded below, so converges to some D.

2. In Proposition 2.4, we use a compactness argument to show that there exists a strictly increasing sequence k_r and a random variable T such that

$$\lim_{r\to\infty} D(T_{k_r}) = D(T).$$

Further,

$$\lim_{r\to\infty} D(T_{k_r+1}) = \lim_{r\to\infty} D\left(\frac{T_{k_r} + T'_{k_r}}{\sqrt{2}}\right) = D\left(\frac{T + T'}{\sqrt{2}}\right),$$

where the T'_{k_r} and T' are independent copies of T_{k_r} and T respectively.

3. We combine these two results: since $D(T_{k_r})$ and $D(T_{k_r+1})$ are both subsequences of the convergent subsequence $D(T_k)$, they must have a common limit. That is,

$$D = D(T) = D\left(\frac{T + T'}{\sqrt{2}}\right),$$

so by the condition for equality in Proposition 2.3, we deduce that $T \sim N(0, \sigma^2)$ and $D = 0$.

4. Proposition 2.3 implies the standard subadditive relation

$$(m + n)D(S_{m+n}) \leqslant mD(S_m) + nD(S_n).$$

Now Theorem 6.6.1 of Hille (1948) implies that $D(S_n)$ converges to $\inf_n D(S_n) = 0$.

The proof of Theorem 1.2 is given in Section 5.

2. Subadditivity of Mallows distance

The Mallows distance and related metrics originated with a transportation problem posed by Monge in 1781 (Rachev 1984; Dudley 1989, pp. 329–330). Kantorovich generalized this problem, and considered the distance obtained by minimising $\mathbb{E}c(X, Y)$, for a general metric c (known as the cost function), over all joint distributions of pairs (X, Y) with fixed marginals. This distance is also known as the Wasserstein metric. Rachev (1984) reviews applications to differential geometry, infinite-dimensional linear programming and information theory, among many others. Mallows (1972) focused on the metric which we have called d_2, while d_1 is sometimes called the Gini index.

In Lemma 2.2 below, we review the existence and uniqueness of the construction which attains the infimum in Definition 1.1, using the concept of a quasi-monotone function.

Definition 2.1. *A function $k : \mathbb{R}^2 \to \mathbb{R}$ induces a signed measure μ_k on $\mathbb{R}^2$ given by*

$$\mu_k\{(x, x'] \times (y, y']\} = k(x, y) + k(x', y') - k(x, y') - k(x', y).$$

We say that k is quasi-monotone if μ_k is a non-negative measure.

The function $k(x, y) = -|x - y|^r$ is quasi-monotone for $r \geqslant 1$, and if $r > 1$ then the measure μ_k is absolutely continuous, with a density which is positive Lebesgue almost everywhere.

Tchen (1980, Corollary 2.1) gives the following result, a two-dimensional version of integration by parts.

Lemma 2.1. *Let $k(x, y)$ be a quasi-monotone function and let $H_1(x, y)$ and $H_2(x, y)$ be distribution functions with the same marginals, where $H_1(x, y) \leqslant H_2(x, y)$ for all x, y. Suppose there exists an H_1- and H_2-integrable function $g(x, y)$, bounded on compact sets, such that $k(x^B, y^B) \leqslant g(x, y)$, where $x^B = (-B) \vee x \wedge B$. Then*

$$\int k(x, y)\mathrm{d}H_2(x, y) - \int k(x, y)\mathrm{d}H_1(x, y) = \int \big\{ H_2^-(x, y) - H_1^-(x, y)\big\}\mathrm{d}\mu_k(x, y).$$

Here $H_i^-(x, y) = \mathbb{P}(X < x, Y < y)$, where (X, Y) have joint distribution function H_i.

Lemma 2.2. *For $r \geqslant 1$, consider the joint distribution of pairs (X, Y) where X and Y have fixed marginals F_X and F_Y, both in $\mathcal{F}_r$. Then*

$$\mathbb{E}|X - Y|^r \geqslant \mathbb{E}|X^* - Y^*|^r, \tag{2}$$

where $X^* = F_X^{-1}(U)$, $Y^* = F_Y^{-1}(U)$ and $U \sim U(0, 1)$. For $r > 1$, equality is attained only if $(X, Y) \sim (X^*, Y^*)$.

Proof. Observe, as in Fréchet (1951), that if the random variables X, Y have fixed marginals F_X and F_Y, then

$$\mathbb{P}(X \leqslant x, Y \leqslant y) \leqslant H_+(x, y), \tag{3}$$

where $H_+(x, y) = \min(F_X(x), F_Y(y))$. This bound is achieved by taking $U \sim U(0, 1)$ and setting $X^* = F_X^{-1}(U)$, $Y^* = F_Y^{-1}(U)$.

Thus, by Lemma 2.1, with $k(x, y) = -|x - y|^r$, for $r \geqslant 1$, and taking $H_1(x, y) = \mathbb{P}(X \leqslant x, Y \leqslant y)$ and $H_2 = H_+$, we deduce that

$$\mathbb{E}|X - Y|^r - \mathbb{E}|X^* - Y^*|^r = \int \{H_+(x, y) - H_1(x, y)\}\mathrm{d}\mu_k(x, y) \geqslant 0,$$

so (X^*, Y^*) achieves the infimum in the definition of the Wasserstein distance.

Finally, since taking $r > 1$ implies that the measure μ_k has a strictly positive density with respect to Lebesgue measure, we can only have equality in (2) if $\mathbb{P}(X \leqslant x, Y \leqslant y) = \min\{F_X(x), F_Y(y)\}$ Lebesgue almost everywhere. But the joint distribution function is right-continuous, so this condition determines the value of $\mathbb{P}(X \leqslant x, Y \leqslant y)$ everywhere. □

Using the construction in Lemma 2.2, Bickel and Freedman (1981) establish that if X_1 and X_2 are independent and Y_1 and Y_2 are independent, then

$$d_2^2(X_1 + X_2, Y_1 + Y_2) \leqslant d_2^2(X_1, Y_1) + d_2^2(X_2, Y_2). \tag{4}$$

Similar subadditive expressions arise in the proof of convergence of Fisher information in Johnson and Barron (2004). By focusing on the case $r = 2$ in Definition 1.1, and by using the theory of L^2 spaces and projections, we establish parallels with the Fisher information argument.

We prove equation (4) below, and further consider the case of equality in this relation. Major (1978, p. 504) gives an equivalent construction to that given in Lemma 2.2. If F_Y is a continuous distribution function, then $F_Y(Y) \sim U(0, 1)$, so we generate $Y^* \sim F_Y$ and take $X^* = F_X^{-1} \circ F_Y(Y^*)$. Recall that if $\mathbb{E}X = \mu$ and $\operatorname{var} X = \sigma^2$, we write $D(X)$ for $d_2^2(X, Z_{\mu,\sigma^2})$.

Proposition 2.3. *If X_1, X_2 are independent, with finite variances $\sigma_1^2, \sigma_2^2 > 0$, then for any $t \in (0, 1)$,*

$$D\left(\sqrt{t}X_1 + \sqrt{1-t}X_2\right) \leqslant tD(X_1) + (1-t)D(X_2),$$

with equality if and only if X_1 and X_2 are normal.

Proof. We consider bounding $D(X_1 + X_2)$ for independent X_1 and X_2 with mean zero, since the general result follows on translation and rescaling.

We generate independent $Y_i^* \sim N(0, \sigma_i^2)$, and take $X_i^* = F_{X_i}^{-1} \circ \Phi_{\sigma_i^2}(Y_i^*) = h_i(Y_i^*)$, say, for $i = 1, 2$. Further, writing $\sigma^2 = \sigma_1^2 + \sigma_2^2$, we define $Y^* = Y_1^* + Y_2^*$ and set $X^* = F_{X_1+X_2}^{-1} \circ \Phi_{\sigma^2}(Y_1^* + Y_2^*) = h(Y_1^* + Y_2^*)$, say. Then

$$\begin{aligned} d_2^2(X_1 + X_2, Y_1 + Y_2) &= \mathbb{E}(X^* - Y^*)^2 \\ &\leqslant \mathbb{E}(X_1^* + X_2^* - Y_1^* - Y_2^*)^2 \\ &= \mathbb{E}(X_1^* - Y_1^*)^2 + \mathbb{E}(X_2^* - Y_2^*)^2 \\ &= d_2^2(X_1, Y_1) + d_2^2(X_2, Y_2). \end{aligned}$$

Equality holds if and only if $(X_1^* + X_2^*, Y_1^* + Y_2^*)$ has the same distribution as (X^*, Y^*). By our construction of $Y^* = Y_1^* + Y_2^*$, this means that $(X_1^* + X_2^*, Y_1^* + Y_2^*)$ has the same distribution as $(X^*, Y_1^* + Y_2^*)$, so $\mathbb{P}\{X_1^* + X_2^* = h(Y_1^* + Y_2^*)\} = \mathbb{P}\{X^* = h(Y_1^* + Y_2^*)\} = 1$. Thus, if equality holds, then

$$h_1(Y_1^*) + h_2(Y_2^*) = h(Y_1^* + Y_2^*) \text{ almost surely.} \tag{5}$$

Brown (1982) and Johnson and Barron (2004), showed that equality holds in (5) if and only if h, h_1, h_2 are linear. In particular, Proposition 2.1 of Johnson and Barron (2004) implies that there exist constants a_i and b_i such that

$$\begin{aligned} &\mathbb{E}\{h(Y_1^* + Y_2^*) - h_1(Y_1^*) - h_2(Y_2^*)\}^2 \\ &\quad \geqslant \frac{2\sigma_1^2\sigma_2^2}{(\sigma_1^2 + \sigma_2^2)^2}\left[\mathbb{E}\{h_1(Y_1^*) - a_1Y_1^* - b_1\}^2 + \mathbb{E}\{h_2(Y_2^*) - a_2Y_2^* - b_2\}^2\right]. \end{aligned} \tag{6}$$

Hence, if (5) holds, then $h_i(u) = a_iu + b_i$ almost everywhere. Since Y_i^* and X_i^* have the same mean and variance, it follows that $a_i = 1$, $b_i = 0$. Hence $h_1(u) = h_2(u) = u$ and $X_i^* = Y_i^*$. □

Recall that $T_k = S_{2^k}$, where $S_n = (X_1 + \ldots + X_n)/\sqrt{n}$ is a normalized sum of independent and identically distributed random variables of mean zero and finite variance σ^2.

Proposition 2.4. *There exists a strictly increasing sequence $(k_r) \in \mathbb{N}$ and a random variable T such that*

$$\lim_{r\to\infty} D(T_{k_r}) = D(T).$$

If T'_{k_r} and T' are independent copies of T_{k_r} and T respectively, then

$$\lim_{r\to\infty} D(T_{k_r+1}) = \lim_{r\to\infty} D\left(\frac{T_{k_r} + T'_{k_r}}{\sqrt{2}}\right) = D\left(\frac{T + T'}{\sqrt{2}}\right).$$

Proof. Since $\mathrm{var}(T_k) = 1$ for all k, the sequence (T_k) is tight. Therefore, by Prohorov's theorem, there exists a strictly increasing sequence (k_r) and a random variable T such that

$$T_{k_r} \xrightarrow{d} T \tag{7}$$

as $r \to \infty$. Moreover, the proof of Lemma 5.2 of Brown (1982) shows that the sequence $(T^2_{k_r})$ is uniformly integrable. But this, combined with (7), implies that $\lim_{r\to\infty} d_2(T_{k_r}, T) = 0$ (Bickel and Freedman 1981, Lemma 8.3(b)). Hence

$$D(T_{k_r}) = d_2^2(T_{k_r}, Z_{\sigma^2}) \leqslant \{d_2(T_{k_r}, T) + d_2(T, Z_{\sigma^2})\}^2 \to d_2^2(T, Z_{\sigma^2}) = D(T)$$

as $r \to \infty$. Similarly, $d_2^2(T, Z_{\sigma^2}) \leqslant \{d_2(T, T_{k_r}) + d_2(T_{k_r}, Z_{\sigma^2})\}^2$, yielding the opposite inequality. This proves the first part of the proposition.

For the second part, it suffices to observe that $T_{k_r} + T'_{k_r} \xrightarrow{d} T + T'$ as $r \to \infty$, and $\mathbb{E}(T_{k_r} + T'_{k_r})^2 \to \mathbb{E}(T + T')^2$, and then use the same argument as in the first part of the proposition. □

Combining Propositions 2.3 and 2.4, as described in Section 1, the proof of Theorem 1.1 is now complete.

3. Convergence of d_r for general r

The subadditive inequality (4) arises in part from a moment inequality; that is, if X_1 and X_2 are independent with mean zero, then $\mathbb{E}|X_1 + X_2|^r \leqslant \mathbb{E}|X_1|^r + \mathbb{E}|X_2|^r$, for $r = 2$. Similar results imply that for $r \geqslant 2$, we have $\lim_{n\to\infty} d_r(S_n, Z_{\sigma^2}) = 0$. First, we prove the following lemma:

Lemma 3.1. *Consider independent random variables $V_1, V_2, \ldots$ and $W_1, W_2, \ldots$, where for some $r \geqslant 2$ and for all i, $\mathbb{E}|V_i|^r < \infty$ and $\mathbb{E}|W_i|^r < \infty$. Then for any m, there exists a constant $c(r)$ such that*

$$d_r^r(V_1 + \ldots + V_m, W_1 + \ldots + W_m) \leqslant c(r)\left\{\sum_{i=1}^m d_r^r(V_i, W_i) + \left(\sum_{i=1}^m d_2^2(V_i, W_i)\right)^{r/2}\right\}.$$

Proof. We consider independent $U_i \sim U(0, 1)$, and set $V_i^* = F_V^{-1}(U_i)$ and $W_i^* = F_W^{-1}(U_i)$. Then

$$\begin{aligned} d_r^r(V_1 + \ldots + V_m, W_1 + \ldots + W_m) &\leqslant \mathbb{E}\left|\sum_{i=1}^m (V_i^* - W_i^*)\right|^r \\ &\leqslant c(r)\left\{\sum_{i=1}^m \mathbb{E}|V_i^* - W_i^*|^r + \left(\sum_{i=1}^m \mathbb{E}|V_i^* - W_i^*|^2\right)^{r/2}\right\} \end{aligned}$$

as required. This final line is an application of Rosenthal's inequality (Petrov 1995, Theorem 2.9) to the sequence $(V_i^* - W_i^*)$. □

Using Lemma 3.1, we establish the following theorem:

Theorem 3.2. *Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean zero, variance $\sigma^2 > 0$ and $\mathbb{E}|X_1|^r < \infty$ for some $r \geqslant 2$. If $S_n = (X_1 + \ldots + X_n)/\sqrt{n}$, then*

$$\lim_{n\to\infty} d_r(S_n, Z_{\sigma^2}) = 0.$$

Proof. Theorem 1.1 covers the case of $r = 2$, so need only consider $r > 2$. We use a scaled version of Lemma 3.1 twice. First, we use $V_i = X_i$, $W_i \sim N(0, \sigma^2)$ and $m = n$, in order to deduce that, by monotonicity of the r-norms,

$$\begin{aligned} d_r^r(S_n, Z_{\sigma^2}) &\leqslant c(r)\left\{n^{1-r/2} d_r^r(X_1, Z_{\sigma^2}) + d_2^2(X_1, Z_{\sigma^2})^{r/2}\right\} \\ &\leqslant c(r)\left(n^{1-r/2} + 1\right) d_r^r(X_1, Z_{\sigma^2}), \end{aligned}$$

so that $d_r^r(S_n, Z_{\sigma^2})$ is uniformly bounded in n, by K, say. Then, for general n, define $N = \lceil\sqrt{n}\rceil$, take $m = \lceil n/N\rceil$, and $u = n - (m-1)N \leqslant N$. In Lemma 3.1, take

$$V_i = X_{(i-1)N+1} + \ldots + X_{iN}, \qquad \text{for } i = 1, \ldots, m-1,$$

$$V_m = X_{(m-1)N+1} + \ldots + X_n,$$

and $W_i \sim N(0, N\sigma^2)$ for $i = 1, \ldots, m-1$, $W_m \sim N(0, u\sigma^2)$ independently.

Now the uniform bound above gives, on rescaling,

$$d_r^r(V_i, W_i) = N^{r/2} d_r^r(S_N, Z_{\sigma^2}) \leqslant N^{r/2} K, \qquad \text{for } i = 1, \ldots m-1,$$

$$d_r^r(V_m, W_m) = u^{r/2} d_r^r(S_u, Z_{\sigma^2}) \leqslant N^{r/2} K.$$

Furthermore, $d_2^2(V_i, W_i) = Nd_2^2(S_N, Z_{\sigma^2})$ for $i = 1, \dots m-1$ and $d_2^2(V_m, W_m) = ud_2^2(S_u, Z_{\sigma^2}) \leqslant Nd_2^2(S_1, Z_{\sigma^2})$. Hence, using Lemma 3.1 again, we obtain

$$\begin{aligned} d_r^r(S_n, Z_{\sigma^2}) &= \frac{1}{n^{r/2}} d_r^r(V_1 + \dots + V_m, W_1 + \dots + W_m) \\ &\leqslant \frac{c(r)}{n^{r/2}} \left\{ \sum_{i=1}^m d_r^r(V_i, W_i) + \left(\sum_{i=1}^m d_2^2(V_i, W_i) \right)^{r/2} \right\} \\ &\leqslant c(r) \left\{ mK \frac{N^{r/2}}{n^{r/2}} + \left(\frac{N(m-1)}{n} d_2^2(S_N, Z_{\sigma^2}) + \frac{N}{n} d_2^2(S_1, Z_{\sigma^2}) \right)^{r/2} \right\} \\ &\leqslant c(r) \left\{ \frac{mK}{(m-1)^{r/2}} + \left(d_2^2(S_N, Z_{\sigma^2}) + \frac{1}{m-1} d_2^2(S_1, Z_{\sigma^2}) \right)^{r/2} \right\}. \end{aligned}$$

This converges to zero since $\lim_{n\to\infty} d_2(S_N, Z_{\sigma^2}) = 0$. □

4. Strengthening subadditivity

Under certain conditions, we obtain a rate for the convergence in Theorem 1.1. Equation (1) shows that $D(T_k)$ is decreasing. Since $D(T_k)$ is bounded below, the difference sequence $D(T_k) - D(T_{k+1})$ converges to zero. As in Johnson and Barron (2004), we examine this difference sequence, to show that its convergence implies convergence of $D(T_k)$ to zero.

Further, in the spirit of Johnson and Barron (2004), we hope that if the difference sequence is small, then equality 'nearly' holds in (5), and so the functions h, h_1, h_2 are 'nearly' linear. This implies that if $\mathrm{cov}(X, Y)$ is close to its maximum, then X will be close to $h(Y)$ in the L^2 sense.

Following del Barrio *et al.* (1999), we define a new distance quantity $D^*(X) = \inf_{m,s^2} d_2^2(X, Z_{m,s^2})$. Notice that $D(X) = 2\sigma^2 - 2\sigma k \leqslant 2\sigma^2$, where $k = \int_0^1 F_X^{-1}(x)\Phi^{-1}(x)\mathrm{d}x$. This follows since F_X^{-1} and Φ^{-1} are increasing functions, so $k \geqslant 0$ by Chebyshev's rearrangement lemma. Using results of del Barrio *et al.* (1999), it follows that

$$D^*(X) = \sigma^2 - k^2 = D(X) - \frac{D(X)^2}{4\sigma^2},$$

and convergence of $D(S_n)$ to zero is equivalent to convergence of $D^*(S_n)$ to zero.

Proposition 4.1. *Let X_1 and X_2 be independent and identically distributed random variables with mean μ, variance $\sigma^2 > 0$ and densities with respect to Lebesgue measure. Defining $g(u) = \Phi_{\mu,\sigma^2}^{-1} \circ F_{(X_1+X_2)/\sqrt{2}}(u)$, if the derivative $g'(u) \geqslant c$ for all u, then*

$$D\left(\frac{X_1 + X_2}{\sqrt{2}}\right) \leqslant \left(1 - \frac{c}{2}\right) D(X_1) + \frac{cD(X_1)^2}{8\sigma^2} \leqslant \left(1 - \frac{c}{4}\right) D(X_1).$$

Proof. As before, translation invariance allows us to take $\mathbb{E}X_i = 0$. For random variables $X, Y,$ we write $g(u) = F_Y^{-1} \circ F_X(u)$, and $h(u) = g^{-1}(u)$. The function $k(x, y) = -\{x - h(y)\}^2$ is quasi-monotone and induces the measure $\mathrm{d}\mu_k(x, y) = 2h'(y)\mathrm{d}x\,\mathrm{d}y$. Taking $H_1(x, y) = \mathbb{P}(X \leqslant x, Y \leqslant y)$ and $H_2(x, y) = \min\{F_X(x), F_Y(y)\}$ in Lemma 2.1 implies that

$$\mathbb{E}\{X - h(Y)\}^2 = 2\int h'(y)\{H_2(x, y) - H_1(x, y)\}\mathrm{d}x\,\mathrm{d}y,$$

since $\mathbb{E}\{X^* - h(Y^*)\}^2 = 0$. By assumption $h'(y) \leqslant 1/c$, so

$$\mathbb{E}\{X - h(Y)\}^2 \leqslant \frac{2}{c}\{\mathrm{cov}(X^*, Y^*) - \mathrm{cov}(X, Y))\}.$$

Again take Y_1^*, Y_2^* independent $N(0, \sigma^2)$ and set $X_i^* = F_{X_i}^{-1} \circ F_{Y_i}(Y_i^*) = h_i(Y_i^*)$. Then define $Y^* = Y_1^* + Y_2^*$ and take $X^* = F_{X_1+X_2}^{-1} \circ F_{Y_1+Y_2}(Y^*)$.
Then there exist a and b such that

$$\begin{aligned}
&d_2^2(X_1, Y_1) + d_2^2(X_2, Y_2) - d_2^2(X_1 + X_2, Y_1 + Y_2)\\
&\qquad = \mathbb{E}(X_1^* + X_2^* - Y_1^* - Y_2^*)^2 - \mathbb{E}(X^* - Y^*)^2\\
&\qquad = 2\mathrm{cov}(X^*, Y^*) - 2\mathrm{cov}(X_1^* + X_2^*, Y_1^* + Y_2^*)\\
&\qquad \geqslant c\mathbb{E}\{X_1^* + X_2^* - h(Y_1^* + Y_2^*)\}^2\\
&\qquad = c\mathbb{E}\{h_1(Y_1^*) + h_2(Y_2^*) - h(Y_1^* + Y_2^*)\}^2\\
&\qquad \geqslant c\mathbb{E}\{h_1(Y_1^*) - aY_1^* - b\}^2 \geqslant cD^*(X_1),
\end{aligned}$$

where the penultimate inequality follows by (6). Recall that $D(X) \leqslant 2\sigma^2$, so that $D^*(X) = D(X) - D(X)^2/(4\sigma^2) \geqslant D(X)/2$. The result follows on rescaling. □

We briefly discuss the strength of the condition imposed. If X has mean zero, distribution function F_X and continuous density f_X, define the scale-invariant quantity

$$\mathcal{C}(X) = \inf_u (\Phi_{\sigma^2}^{-1} \circ F_X)'(u) = \inf_{p\in(0,1)} \frac{f_X(F_X^{-1}(p))}{\phi_{\sigma^2}(\Phi_{\sigma^2}^{-1}(p))} = \inf_{p\in(0,1)} \sigma\frac{f_X(F_X^{-1}(p))}{\phi(\Phi^{-1}(p))}.$$

We want to understand when $\mathcal{C}(X) > 0$.

Example 4.1. If $X \sim U(0, 1)$, then $\mathcal{C}(X) = 1/\sqrt{12\sup_x \phi(x)} = \sqrt{\pi/6}$.

Lemma 4.2. *If X has mean zero and variance σ^2 then $\mathcal{C}(X)^2 \leqslant \sigma^2/(\sigma^2 + \mathrm{median}(X)^2)$.*

Proof. By the mean value inequality, for all p,

$$|\Phi_{\sigma^2}^{-1}(p)| = |\Phi_{\sigma^2}^{-1}(p) - \Phi_{\sigma^2}^{-1}(1/2)| \geqslant \mathcal{C}(X)|F_X^{-1}(p) - F_X^{-1}(1/2)|,$$

so that

$$\sigma^2 + F_X^{-1}(1/2)^2 = \int_0^1 F_X^{-1}(p)^2 \mathrm{d}p + F_X^{-1}(1/2)^2 = \int_0^1 \{F_X^{-1}(p) - F_X^{-1}(1/2)\}^2 \mathrm{d}p$$
$$\leqslant \frac{1}{\mathcal{C}(X)^2} \int_0^1 \Phi_{\sigma^2}^{-1}(p)^2 \mathrm{d}p = \frac{\sigma^2}{\mathcal{C}(X)^2}.$$

□

In general, we are concerned with the rate at which $f_X(x) \to 0$ at the edges of the support.

Lemma 4.3. *If for some* $\epsilon > 0$ *and* $c > 0$,

$$f_X(F_X^{-1}(p)) \geqslant c(1-p)^{1-\epsilon}, \qquad \text{for } p \text{ close to } 1, \tag{8}$$

then $\lim_{p \to 1} f_X(F_X^{-1}(p))/\phi(\Phi^{-1}(p)) = \infty$. *Correspondingly, if*

$$f_X(F_X^{-1}(p)) \geqslant c p^{1-\epsilon}, \qquad \text{for } p \text{ close to } 0, \tag{9}$$

then $\lim_{p \to 0} f_X(F_X^{-1}(p))/\phi(\Phi^{-1}(p)) = \infty$.

Proof. Simply note that by the Mills ratio (Shorack and Wellner 1986, p. 850) as $x \to \infty$, $\Phi(x) \sim \phi(x)/x$, so that as $p \to 1$, $\phi(\Phi^{-1}(p)) \sim (1-p)\Phi^{-1}(p) \sim (1-p)\sqrt{-2\log(1-p)}$.

□

Example 4.2. (i) The density of the n-fold convolution of $U(0, 1)$ random variables is given by $f_X(x) = x^{n-1}/(n-1)!$ for $0 < x < 1$, hence $F_X^{-1}(p) = (n!p)^{1/n}$ for p sufficiently small, and $f_X(F_X^{-1}(p)) = n/(n!)^{1/n} p^{(n-1)/n}$, so that equation (9) holds.

(ii) For an Exp(1) random variable, $f_X(F_X^{-1}(p)) = 1 - p$, so that equation (8) fails and $\mathcal{C}(X) = 0$.

To obtain bounds on $D(S_n)$ as $n \to \infty$, we need to control the sequence $\mathcal{C}(S_n)$. Motivated by properties of the (seemingly related) Poincaré constant, we conjecture that $\mathcal{C}((X_1 + X_2)/\sqrt{2}) \geqslant \mathcal{C}(X_1)$ for independent and identically distributed X_i. If this is true and $\mathcal{C}(X) = c$ then $\mathcal{C}(S_n) \geqslant c$ for all n.

Assuming that $\mathcal{C}(S_n) \geqslant c$ for all n, note that $D(T_k) \leqslant (1 - c/4)^k D(X_1) \leqslant (1 - c/4)^k (2\sigma^2)$. Now

$$D(T_{k+1}) \leqslant D(T_k)(1 - c/2)\left\{1 + \frac{cD(T_k)}{8\sigma^2(1 - c/2)}\right\},$$

so

$$\prod_{k=0}^{\infty}\left\{1 + \frac{cD(T_k)}{8\sigma^2(1 - c/2)}\right\} \leqslant \exp\left\{\sum_{k=0}^{\infty} \frac{cD(T_k)}{8\sigma^2(1 - c/2)}\right\} \leqslant \exp\left(\frac{1}{1 - c/2}\right).$$

We deduce that

$$D(T_k) \leqslant D(X_1)\exp\left(\frac{1}{1-c/2}\right)(1-c/2)^k,$$

or $D(S_n) = O(n^t)$, where $t = \log_2(1-c/2)$.

Remark 4.1. In general, convergence of $d_4(S_n, Z_{\sigma^2})$ cannot occur at a rate faster than $O(1/n)$. This follows because $\mathbb{E}S_n^4 = 3\sigma^4 + \gamma(X_1)/n$, where $\gamma(X)$, the excess kurtosis, is defined by $\gamma(X) = \mathbb{E}X^4 - 3(\mathbb{E}X^2)^2$ (when $\mathbb{E}X = 0$). Thus by Minkowski's inequality,

$$\begin{aligned} d_4(S_n, Z_{\sigma^2}) &\geqslant |(\mathbb{E}S_n^4)^{1/4} - (\mathbb{E}Z_{\sigma^2}^4)^{1/4}| \\ &= 3^{1/4}\sigma\left|\left(1+\frac{\gamma(X)}{n}\right)^{1/4} - 1\right| = \frac{3^{1/4}\sigma|\gamma(X)|}{4n} + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Motivated by this remark, and by analogy with the rates discovered in Johnson and Barron (2004), we conjecture that the true rate of convergence is $D(S_n) = O(1/n)$. To obtain this, we would need to control $1 - \mathcal{C}(S_n)$.

5. Convergence to stable distributions

We now consider convergence to other stable distributions. Gnedenko and Kolmogorov (1954) review classical results of this kind. We say that Y is α-stable if, when $Y_1, \ldots Y_n$ are independent copies of Y, we have $(Y_1 + \ldots + Y_n - b_n)/n^{1/\alpha} \sim Y$ for some sequence (b_n). Note that α-stable variables only exist for $0 < \alpha \leqslant 2$; we assume for the rest of this section that $\alpha < 2$.

Definition 5.1. *If X has a distribution function of the form*

$$F_X(x) = \frac{c_1 + b_X(x)}{|x|^\alpha}, \qquad \text{for } x < 0,$$

$$1 - F_X(x) = \frac{c_2 + b_X(x)}{x^\alpha}, \qquad \text{for } x \geqslant 0,$$

where $b_X(x) \to 0$ as $x \to \pm\infty$, then we say that X is in the domain of normal attraction of some stable Y with tail parameters c_1, c_2.

Theorem 5 of Section 35 of Gnedenko and Kolmogorov (1954) shows that if F_X is of this form, there exist a sequence (a_n) and an α-stable distribution function F_Y, determined by the parameters α, c_1, c_2, such that

$$\frac{X_1 + \ldots + X_n - a_n}{n^{1/\alpha}} \xrightarrow{d} F_Y. \tag{10}$$

Although (10) is obviously very similar to the standard central limit theorem, one important distinguishing feature is that both $\mathbb{E}|X|^\alpha$ and $\mathbb{E}|Y|^\alpha$ are infinite for $0 < \alpha < 2$.

We use the following moment bounds from von Bahr and Esseen (1965). If $X_1, X_2, \ldots$ are independent, then

$$\mathbb{E}|X_1 + \ldots + X_n|^r \leqslant \sum_{i=1}^{n} \mathbb{E}|X_i|^r, \qquad \text{for } 0 < r \leqslant 1, \tag{11}$$

$$\mathbb{E}|X_1 + \ldots + X_n|^r \leqslant 2\sum_{i=1}^{n} \mathbb{E}|X_i|^r, \qquad \text{when } \mathbb{E}X_i = 0, \text{ for } 1 < r \leqslant 2. \tag{12}$$

Now, using ideas of Stout (1979), we show that for a subset of the domain of normal attraction, $d_\beta(X, Y) < \infty$, for some $\beta > \alpha$.

Definition 5.2. *We say that a random variable is in the domain of strong normal attraction of Y if the function $b_X(x)$ from Definition 5.1 satisfies*

$$b_X(x) \leqslant \frac{C}{|x|^\gamma},$$

for some constant C and some $\gamma > 0$.

Cramér (1963) shows that such random variables have an Edgeworth-style expansion, and thus convergence to Y occurs. However, his proof requires some involved analysis and use of characteristic functions. See also Mijnheer (1984, 1986), which use bounds based on the quantile transformation described above.

We can regard Definition 5.2 as being analogous to requiring a bounded $(2+\delta)$th moment in the central limit theorem, which allows an explicit rate of convergence (via the Berry–Esseen theorem). We now show the relevance of Definition 5.2 to the problem of stable convergence.

Lemma 5.1. *If X is in the domain of strong normal attraction of an α-stable random variable Y, then $d_\beta(X, Y) < \infty$ for some $\beta > \alpha$.*

Proof. We show that Major's construction always gives a joint distribution (X^*, W^*) with $\mathbb{E}|X^* - W^*|^\beta < \infty$, and hence $d_\beta(X, W) < \infty$. Following Stout (1979), define a random variable W by

$$\mathbb{P}(W \geqslant x) = c_2 x^{-\alpha}, \qquad \text{if } x > (2c_2)^{1/\alpha},$$

$$\mathbb{P}(W \leqslant x) = c_1 |x|^{-\alpha}, \qquad \text{if } x < -(2c_1)^{1/\alpha},$$

$$\mathbb{P}(W \in [-(2c_1)^{1/\alpha}, (2c_2)^{1/\alpha}]) = 0.$$

Then for $w > 1/2$, $F_W^{-1}(w) = \{c_2/(1-w)\}^{1/\alpha}$, and so for $x \geqslant 0$,

$$x - F_W^{-1}(F_X(x)) = x\left\{1 - \left(\frac{c_2}{c_2 + b_X(x)}\right)^{1/\alpha}\right\}.$$

Now, since $b_X(x) \to 0$, there exists K such that if $x \geqslant K$ then $b_X(x) \geqslant -c_2/2$.

By the mean value inequality, if $t \geqslant -1/2$, then

$$|1 - (1+t)^{-1/\alpha}| \leqslant \frac{|t|2^{1+1/\alpha}}{\alpha},$$

so that for $x \geqslant K$,

$$|x - F_W^{-1} F_X(x)| \leqslant \frac{2^{1+1/\alpha} x |b_X(x)|}{\alpha c_2}.$$

Thus, if X is in the strong domain of attraction, then

$$\int_{|x| \geqslant K} \left| x - F_W^{-1} F_X(x) \right|^\beta \mathrm{d}F_X(x) \leqslant \left(\frac{2^{1+1/\alpha} C}{\alpha c_2} \right)^\beta \int_{|x| \geqslant K} |x|^{\beta(1-\gamma)} \mathrm{d}F_X(x).$$

Hence $d_\beta(X, W)$ is finite for all β if $\gamma \geqslant 1$ and for $\beta < \alpha/(1-\gamma)$ if $\gamma < 1$. Moreover, Mijnheer (1986, equation (2.2)) shows that if Y is α-stable, then as $x \to \infty$,

$$\mathbb{P}(Y \geqslant x) = \frac{c_2}{x^\alpha} + O\left(\frac{1}{x^{2\alpha}} \right),$$

and so Y is in its own domain of strong normal attraction. Thus using the construction above, $d_\beta(Y, W)$ is finite for all β if $\alpha \geqslant 1$ and for $\beta < \alpha/(1-\alpha)$ otherwise.

Recall that the triangle inequality holds, for d_β or d_β^β, according to whether $\beta \geqslant 1$ or $\beta < 1$. Hence, $d_\beta(X, Y)$ is finite for all β if $\min(\alpha, \gamma) \geqslant 1$ and for $\beta < \alpha/(1 - \min(\alpha, \gamma))$ otherwise. □

Note that for random variables X_i in the same strong domain of normal attraction, $d_\beta(X_i, Y)$ may be bounded in terms of the function $b_{X_i}(x)$. In particular, if there exist C, γ such that $b_{X_i}(x) \leqslant C/|x|^\gamma$ then $\sup_i d_\beta(X_i, Y) < \infty$, so the hypothesis of Theorem 1.2 is satisfied.

Proof of Theorem 1.2. We use the bounds provided by (11) and (12). We consider independent pairs (X_i^*, Y_i^*) having the joint distribution that achieves the infimum in Definition 1.1. Then by rescaling we have that

$$\begin{aligned} d_\beta^\beta(S_n, Y) &\leqslant \frac{1}{n^{\beta/\alpha}} d_\beta^\beta(X_1 + \ldots + X_n, Y_1 + \ldots + Y_n) \\ &\leqslant \frac{1}{n^{\beta/\alpha}} \mathbb{E} \left| \sum_{i=1}^n (X_i^* - Y_i^*) \right|^\beta \leqslant \frac{2}{n^{\beta/\alpha}} \sum_{i=1}^n \mathbb{E}|X_i^* - Y_i^*|^\beta. \end{aligned}$$

We deduce that in the case of identical variables, $d_\beta(S_n, Y)$ (and hence $d_\alpha(S_n, Y)$) converges at rate $O(n^{1/\beta - 1/\alpha})$. □

We now combine Theorem 1.2 and Lemma 5.1, to obtain a rate of convergence for identical variables. Note that Theorem 1.2 requires us to take $\beta \leqslant 2$. Overall, then, we deduce that $d_\alpha(S_n, Y)$ converges at rate $O(n^{-t})$, where

1. if $\min(\alpha, \gamma) \geqslant 1$, we take $\beta = 2$, and hence $t = 1/\alpha - 1/2$;
2. if $\min(\alpha, \gamma) < 1$, we may take $\beta = \min[\alpha/\{1 - \min(\alpha, \gamma) + \epsilon\}, 2]$ for any $\epsilon > 0$, and then $t = \min(1/\alpha - 1/2, 1 - \epsilon, \gamma/\alpha - \epsilon)$.

Theorem 3.2 implies that if $d_r(S_n, Z_{\sigma^2})$ ever becomes finite, then it tends to zero, the counterpart of the following result, the proof of which was suggested by an anonymous referee.

Theorem 5.2. *Fix $\alpha \in (0, 2)$, let $X_1, X_2, \ldots$ be independent random variables (where $\mathbb{E}X_i = 0$ if $\alpha > 1$), and let $S_n = (X_1 + \ldots + X_n)/n^{1/\alpha}$. Suppose there exist an α-stable random variable Y and $Y_1, Y_2, \ldots$ having the same distribution as Y and satisfying*

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\{|X_i - Y_i|^\alpha \mathbb{I}(|X_i - Y_i| > b)\} \to 0 \qquad \text{as } b \to \infty. \tag{13}$$

If $\alpha \neq 1$ then $\lim_{n\to\infty} d_\alpha(S_n, Y) = 0$, and if $\alpha = 1$ then there exists a sequence $c_n = n^{-1}\sum_{i=1}^{n}\mathbb{E}(X_i - Y_i)$ such that $\lim_{n\to\infty} d_\alpha(S_n - c_n, Y) = 0$.

Proof. Fix $\epsilon > 0$. Suppose first that $1 \leqslant \alpha < 2$ and let $d_i = \mathbb{E}(X_i - Y_i)$. Note that $d_i = 0$ if $\alpha > 1$. Let $b > 0$ and define

$$U_i = (X_i - Y_i)\mathbb{I}(|X_i - Y_i| \leqslant b) - \mathbb{E}\{(X_i - Y_i)\mathbb{I}(|X_i - Y_i| \leqslant b)\},$$
$$V_i = (X_i - Y_i)\mathbb{I}(|X_i - Y_i| > b) - \mathbb{E}\{(X_i - Y_i)\mathbb{I}(|X_i - Y_i| > b)\}.$$

Then by (12),

$$\begin{aligned}
d_\alpha^\alpha(S_n - c_n, Y) &\leqslant \frac{1}{n}\mathbb{E}\left|\sum_{i=1}^n (X_i - Y_i - d_i)\right|^\alpha = \frac{1}{n}\mathbb{E}\left|\sum_{i=1}^n U_i + \sum_{i=1}^n V_i\right|^\alpha \\
&\leqslant \frac{2^{\alpha-1}}{n}\mathbb{E}\left|\sum_{i=1}^n U_i\right|^\alpha + \frac{2^{\alpha-1}}{n}\mathbb{E}\left|\sum_{i=1}^n V_i\right|^\alpha \\
&\leqslant \frac{2^{\alpha-1}}{n}\left\{\mathbb{E}\left(\sum_{i=1}^n U_i\right)^2\right\}^{\alpha/2} + \frac{2^\alpha}{n}\sum_{i=1}^n \mathbb{E}|V_i|^\alpha \\
&\leqslant \frac{2^{\alpha-1}}{n}\left(\sum_{i=1}^n \mathbb{E}U_i^2\right)^{\alpha/2} + \frac{2^{2\alpha-1}}{n}\sum_{i=1}^n \mathbb{E}\{|X_i - Y_i|^\alpha \mathbb{I}(|X_i - Y_i| > b)\} \\
&\quad + \frac{2^{2\alpha-1}}{n}\sum_{i=1}^n [\mathbb{E}\{|X_i - Y_i|\mathbb{I}(|X_i - Y_i| > b)\}]^\alpha \\
&\leqslant \frac{2^{\alpha-1}b^\alpha}{n^{1-\alpha/2}} + \frac{2^{2\alpha}}{n}\sum_{i=1}^n \mathbb{E}\{|X_i - Y_i|^\alpha \mathbb{I}(|X_i - Y_i| > b)\}.
\end{aligned}$$

The result follows on choosing b sufficiently large to control the second term, and then n sufficiently large to control the first.

For $0 < \alpha < 1$, take U_i as before, $V_i = (X_i - Y_i)\mathbb{I}(|X_i - Y_i| > b)$ and $a_i = \mathbb{E}\{(X_i - Y_i)\mathbb{I}(|X_i - Y_i| \leqslant b)\}$. Now using (11),

$$\begin{aligned}
d_\alpha^\alpha(S_n, Y) &\leqslant \frac{1}{n}\mathbb{E}\left|\sum_{i=1}^n U_i + \sum_{i=1}^n V_i + \sum_{i=1}^n a_i\right|^\alpha \\
&\leqslant \frac{1}{n}\mathbb{E}\left|\sum_{i=1}^n U_i\right|^\alpha + \frac{1}{n}\mathbb{E}\left|\sum_{i=1}^n V_i\right|^\alpha + \frac{1}{n}\left|\sum_{i=1}^n a_i\right|^\alpha \\
&\leqslant \frac{1}{n}\left\{\mathbb{E}\left(\sum_{i=1}^n U_i\right)^2\right\}^{\alpha/2} + \frac{1}{n}\sum_{i=1}^n \mathbb{E}|V_i|^\alpha + \frac{b^\alpha}{n^{1-\alpha}},
\end{aligned}$$

so again, since b is arbitrary, the result follows. □

Note that when $X_1, X_2, \ldots$ are identically distributed, the Lindeberg condition (13) reduces to the requirement that $d_\alpha(X_1, Y) < \infty$.

References

Barron, A.R. (1986) Entropy and the central limit theorem. *Ann. Probab.*, **14**, 336–342.

Bickel, P.J. and Freedman, D.A. (1981) Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**, 1196–1217.

Brown, L.D. (1982) A proof of the central limit theorem motivated by the Cramér–Rao inequality. In G. Kallianpur, P.R. Krishnaiah, and J.K. Ghosh (eds), *Statistics and Probability, Essays in Honour of C.R. Rao*, pp. 141–148. Amsterdam: North-Holland.

Cramér, H. (1963) On asymptotic expansions for sums of independent random variables with a limiting stable distribution. *Sankhyā Ser. A*, **25**, 13–24.

Csiszár, I. (1965) A note on limiting distributions on topological groups. *Publ. Math. Inst. Hungar. Acad. Sci. Ser. A*, **9**, 595–599.

del Barrio, E., Cuesta-Albertos, J.A., Matrán, C. and Rodríguez-Rodríguez, J.M. (1999) Tests of goodness of fit based on the L_2-Wasserstein distance. *Ann. Statist.*, **27**, 1230–1239.

Dudley, R.M. (1989) *Real Analysis and Probability.* Pacific Grove, CA: Wadsworth and Brooks/Cole Advanced Books and Software.

Fréchet, M. (1951) Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon. Sect. A. (3)*, **14**, 53–77.

Gnedenko, B.V. and Kolmogorov, A.N. (1954) *Limit Distributions for Sums of Independent Random Variables*. Cambridge, MA: Addison-Wesley.

Hille, E. (1948) *Functional Analysis and Semi-groups*, Amer. Math. Soc. Colloq. Publ. 31. New York: American Mathematical Society.

Johnson, O.T. and Barron, A.R. (2004) Fisher information inequalities and the central limit theorem. *Probab. Theory Related Fields*, **129**, 391–409.

Kendall, D.G. (1963) Information theory and the limit theorem for Markov chains and processes with a countable infinity of states. *Ann. Inst. Statist. Math.*, **15**, 137–143.

Major, P. (1978) On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.*, **8**, 487–517.

Mallows, C.L. (1972) A note on asymptotic joint normality. *Ann. Math. Statist.*, **43**, 508–515.

Mijnheer, J. (1984) On the rate of convergence to a stable limit law. In P. Révész (ed.), *Limit theorems in Probability and Statistics*, Colloq. Math. Soc. János Bolyai 36, Vol. 2, pp. 761–775. Amsterdam: North-Holland.

Mijnheer, J. (1986) On the rate of convergence to a stable limit law. II. *Litovsk. Mat. Sb.*, **26**, 482–487.

Petrov, V.V. (1995) *Limit Theorems of Probability Theory. Sequences of Independent Random Variables*. Oxford: Clarendon Press.

Prohorov, Y. (1952) On a local limit theorem for densities. *Dok. Akad. Nauk SSSR (N.S.)* **83**, 797–800. In Russian.

Rachev, S.T. (1984) The Monge–Kantorovich problem on mass transfer and its applications in stochastics. *Teor. Veroyatnost. i Primenen.*, **29**, 625–653.

Rachev, S.T. and Rüschendorf, L. (1992) Rate of convergence for sums and maxima and doubly ideal metrics. *Teor. Veroyatnost. i Primenen.*, **37**, 276–289.

Rachev, S.T. and Rüschendorf, L. (1994) On the rate of convergence in the CLT with respect to the Kantorovich metric, In *Probability in Banach spaces, 9 (Sandjberg, 1993)*, Progr. Probab. 35, pp. 193–207. Boston: Birkhäuser.

Rényi, A. (1961) On measures of entropy and information. In J. Neyman (ed.), *Proceedings of the Fourth Berkeley Conference on Mathematical Statistics and Probability*, pp. 547–561. Berkeley: University of California Press.

Shimizu, R. (1975) On Fisher's amount of information for location family. In G.P. Patil, S. Kotz and J.K. Ord (eds), *A Modern Course on Statistical Distributions in Scientific Work, Vol. 3*, pp. 305–312. Dordrecht: Reidel.

Shorack, G.R. and Wellner J.A. (1986) *Empirical Processes with Applications to Statistics*. New York: Wiley.

Stout, W. (1979) Almost sure invariance principles when $EX_2^1 = \infty$. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **49**, 23–32.

Tchen, A.H. (1980) Inequalities for distributions with given marginals. *Ann. Probab.*, **8**, 814–827.

von Bahr, B. and Esseen, C.-G. (1965) Inequalities for the *r*th absolute moment of a sum of independent random variables, $1 \leqslant r \leqslant 2$. *Ann. Math. Statist.*, **36**, 299–303.

Received June 2004 and revised February 2005