Math-Net.Ru

Общероссийский математический портал

T. Cacoullos, N. Papadatos, V. Papathanasiou, An Application of a Density Transform and the Local Limit Theorem, *Теория вероятн. и ее примен.*, 2001, том 46, выпуск 4, 803–810

DOI: <https://doi.org/10.4213/tvp3828>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.172.240.79

26 апреля 2023 г., 17:57:41

СПИСОК ЛИТЕРАТУРЫ

1. Kesten H. Sums of independent random variables — without moment conditions. — Ann. Math. Statist., 1972, v. 43, № 3, p. 702–732.
2. Хенгартнер В., Теодореску Р. Функции концентрации. М.: Наука, 1980, 173 с.

Поступила в редакцию
29.I.2001

© 2001 г.

CACOULOS T.*, PAPADATOS N.**,
PAPATHANASIOU V.¹⁾

AN APPLICATION OF A DENSITY TRANSFORM AND THE LOCAL LIMIT THEOREM²⁾

Рассмотрим абсолютно непрерывную случайную величину X с конечной дисперсией σ^2 . Известно, что существует другая случайная величина X^* (которую можно рассматривать как преобразование X) с унимодальной плотностью, удовлетворяющая обобщенному ковариационному тождеству типа Стейна $\text{Cov}[X, g(X)] = \sigma^2 \mathbf{E}[g'(X^*)]$ для любой абсолютно непрерывной функции g с производной g' , такой, что $\mathbf{E}|g'(X^*)| < \infty$. С помощью этого преобразования получены верхние границы для расстояния по вариации между двумя абсолютно непрерывными случайными величинами X и Y . В качестве приложения предлагается доказательство локальной предельной теоремы для сумм независимых одинаково распределенных случайных величин в наиболее общей постановке.

Ключевые слова и фразы: преобразование плотности, локальная предельная теорема для плотностей, полная вариация.

1. Introduction. For an absolutely continuous (a.c.) random variable (r.v.) X with density f , mean μ , variance σ^2 , and support $S(X)$ (through this paper we will always mean the support of an absolutely continuous r.v. X with density f to be the set $S(X) = \{x: f(x) > 0\}$), consider the r.v. X^* with density f^* given by the relation

$$f^*(x) = \frac{1}{\sigma^2} \int_{-\infty}^x (\mu - t)f(t) dt = \frac{1}{\sigma^2} \int_x^{\infty} (t - \mu)f(t) dt. \quad (1.1)$$

The following known properties of X^* (see [9] or [15]) will be used in the sequel.

Lemma 1.1. (i) *The density f^* is unimodal and absolutely continuous with mode μ and maximal value $f^*(\mu) = \mathbf{E}|X - \mu|/(2\sigma^2)$.*

(ii) *For any absolutely continuous function g such that $\mathbf{E}|g'(X^*)| < \infty$,*

$$\text{Cov}[X, g(X)] = \sigma^2 \mathbf{E}[g'(X^*)]. \quad (1.2)$$

Moreover, X^ is characterized by this property.*

* University of Athens, Department of Mathematics, Panepistemiopolis, 157 84 Athens, Greece.

** University of Cyprus, Department of Mathematics and Statistics, P. O. Box 20537, Nicosia 1678, Cyprus.

¹⁾ Part of this work was done when the author was visiting the University of Cyprus.

²⁾ Research partially supported by the Ministry of Industry, Energy and Technology of Greece, grant 1369.

(iii) $S(X^*) = (\text{ess inf } S(X), \text{ess sup } S(X))$.

(iv) For any real numbers $a \neq 0$ and b , $(aX + b)^* \stackrel{d}{=} aX^* + b$ (through this paper, $X \stackrel{d}{=} Y$ will always means that the r.v.'s X and Y have the same distribution).

(v) If the independent a.c. r.v.'s X_1, X_2 have means μ_1, μ_2 and variances σ_1^2, σ_2^2 , then for all a_1 and a_2 with $a_1 a_2 \neq 0$,

$$(a_1 X_1 + a_2 X_2)^* \stackrel{d}{=} B(a_1 X_1^* + a_2 X_2^*) + (1 - B)(a_1 X_1 + a_2 X_2^*), \quad (1.3)$$

where X_1, X_2, X_1^*, X_2^* , and B are mutually independent with $\mathbf{P}\{B = 1\} = a_1^2 \sigma_1^2 / (a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2) = 1 - \mathbf{P}\{B = 0\}$.

Identity (1.2) generalizes the well-known Stein identity for the standard normal r.v. Z , namely $\mathbf{E}[Zg(Z)] = \mathbf{E}[g'(Z)]$, for any a.c. function g with derivative g' satisfying $\mathbf{E}|g'(Z)| < \infty$ (see [18], [17]). The classical Stein identity has had many interesting applications to several areas of probability and statistics (see, e.g., [11] and [12]). Other important applications include the well-known Stein method for the approximation of the distribution of a sum of dependent r.v.'s ([17]). Another point of view is given by Chen ([7]), where the CLT is obtained by using Poincaré type inequalities (closely connected with this identity), and the results of [4].

Some applications of (1.2), concerning variance bounds and characterizations of distributions, are discussed in [15]. Moreover, Goldstein and Reinert ([9]) extended Stein's method and they presented very interesting applications of (1.2) in the rate of convergence in the CLT. In particular, they obtained an $O(n^{-1})$ bound for smooth functions (see [9, Corollary 3.1]). Furthermore, they used the so-called *zero biased transformation* of X (i.e., the r.v. X^*), in order to obtain estimates for the rate of convergence for several dependent samples.

The purpose of the present paper is in obtaining general upper bounds for the total variation distance between two arbitrary a.c. r.v.'s X and Y with finite second moment, provided that $S(X^*) \subset S(Y)$ and that $S(Y)$ is a (finite or infinite) interval (in fact, this does not impose any further assumption if $S(Y) = (-\infty, \infty)$). The results of [14] are extended in two directions: (i) it does not required an interval support on X and (ii) without any further assumptions, closed forms for the constants appearing in the bounds are derived (in terms of the «limiting» r.v. Y).

In the most interesting special case where Y is normal, the bound becomes extremely simple and useful. It is used not only to characterize the normal distribution as a unique fixed point of the zero bias transformation (see [9, Lemma 2.1]), but also to obtain the corresponding stability result (Theorem 2.2). As a final application, an elementary and relatively simple proof of the local limit theorem, due to Prokhorov [16], for sums of i.i.d. r.v.'s, is given in Section 3.

2. Total variation distance. The present section deals with the derivation of upper bounds for the total variation distance

$$d_{TV}(X, Y) = \sup_A \left\{ |\mathbf{P}\{X \in A\} - \mathbf{P}\{Y \in A\}|, A \text{ Borel} \right\} = \frac{1}{2} \int_{-\infty}^{\infty} |f(x) - h(x)| dx,$$

between two r.v.'s X, Y with densities f, h , means μ, m and variances σ^2, s^2 , respectively. We may view X as an arbitrary r.v. and Y as the «limiting» one. The results apply to the case, where $S(Y)$ is an open (finite or infinite) interval and $S(X^*) \subset S(Y)$, i.e., when the measure produced by X is a.c. with respect to that produced by Y (this does not impose any restriction if $S(Y) = (-\infty, \infty)$).

The results generalize (see [14]) for an arbitrary X (not necessarily having an interval support). Furthermore, without using any further assumptions, we derive closed forms for the constants appearing in the bounds in terms of the «limiting» r.v. Y .

Assume that $S(Y) = (a, b)$, where $-\infty \leq a < b \leq \infty$. Fix an arbitrary Borel set A and consider the a.c. function (c.f. [3], [6], [14])

$$\psi_A(x) = \frac{1}{h^*(x)} \int_a^x \left(I(t \in A) - \mathbf{P}\{Y \in A\} \right) h(t) dt, \quad a < x < b, \quad (2.1)$$

where h^* is the density of Y^* . The following lemma is required for the main result.

Lemma 2.1. *The functions $\psi_A(x)$ and $h^*(x)\psi'_A(x)/h(x)$ (the second one defined for almost all $x \in (a, b)$) are absolutely bounded by some finite constants c and c' , respectively, which do not depend on $x \in (a, b)$ and A . A possible choice is $c' = 2$ and $c = s^2 c_Y$, where*

$$c_Y = \frac{2 \max\{\mathbf{P}\{Y \leq m\}, \mathbf{P}\{Y \geq m\}\}}{\mathbf{E}|Y - m|} < \frac{2}{\mathbf{E}|Y - m|}.$$

P r o o f. Let $a < x < x + y \leq m$. It follows that

$$\begin{aligned} \int_a^x h(t) dt \int_x^{x+y} (m-t) h(t) dt &\leq (m-x) \int_a^x h(t) dt \int_x^{x+y} h(t) dt \\ &\leq \int_a^x (m-t) h(t) dt \int_x^{x+y} h(t) dt. \end{aligned}$$

By adding to both sides the quantity $\int_a^x h(t) dt \int_a^x (m-t) h(t) dt$ we conclude that $s^2 h^*(x) H(x+y) \geq s^2 h^*(x+y) H(x)$, where H is the distribution function of Y . Since $x > a$, $h^*(x) > 0$, and $H(x) > 0$ and it follows that the function $H(x)/h^*(x)$ is non-decreasing for $x \in (a, m]$. Similarly, $(1 - H(x))/h^*(x)$ is nonincreasing for $x \in [m, b)$. Since

$$|\psi_A(x)| \leq \frac{1}{h^*(x)} \min\{H(x), 1 - H(x)\}$$

for all $x \in (a, b)$, it follows that

$$|\psi_A(x)| \leq \frac{1}{h^*(m)} \max\{H(m), 1 - H(m)\} = \frac{2s^2 \max\{H(m), 1 - H(m)\}}{\mathbf{E}|Y - m|}$$

by Lemma 1.1(i). Taking derivatives in (2.1) we have that, for almost all $x \in (a, b)$,

$$\frac{h^*(x)}{h(x)} \psi'_A(x) = \frac{x-m}{s^2} \psi_A(x) + (I(x \in A) - \mathbf{P}\{Y \in A\}). \quad (2.2)$$

Hence,

$$\left| \frac{h^*(x)}{h(x)} \psi'_A(x) \right| \leq \frac{1}{s^2} |(x-m)\psi_A(x)| + 1 \leq \frac{|x-m| \min\{H(x), 1 - H(x)\}}{s^2 h^*(x)} + 1.$$

Therefore, for almost all $x \in (a, m]$,

$$\left| \frac{h^*(x)}{h(x)} \psi'_A(x) \right| \leq \frac{(m-x) \int_a^x h(t) dt}{\int_a^x (m-t) h(t) dt} + 1 \leq 2. \quad (2.3)$$

Similarly, $|h^*(x)\psi'_A(x)/h(x)| \leq 2$ for almost all $x \in [m, b)$, and the proof is complete.

The following result is now an immediate consequence of the above lemma.

Theorem 2.1. *For X and Y as above,*

$$d_{TV}(X, Y) \leq 2 \int_a^b \left| f(x) - \frac{\sigma^2 h(x)}{s^2 h^*(x)} f^*(x) \right| dx + c_Y |\mu - m|, \quad (2.4)$$

where f^* is the density of X^* .

P r o o f. Since $f(x) = f(x)I(x \in S(X^*))$ almost everywhere, we may assume that $S(X) \subset S(X^*)$. Taking expectations with respect to X in (2.2) and using Lemma 1.1(ii) we have (c.f. [14])

$$\begin{aligned} \mathbf{P}\{X \in A\} - \mathbf{P}\{Y \in A\} &= \mathbf{E} \left[\frac{h^*(X)}{h(X)} \psi'_A(X) \right] - \frac{1}{s^2} \text{Cov}[X, \psi_A(X)] - \frac{\mu - m}{s^2} \mathbf{E}[\psi_A(X)] \\ &= \mathbf{E} \left[\frac{h^*(X)}{h(X)} \psi'_A(X) \right] - \frac{\sigma^2}{s^2} \mathbf{E}[\psi'_A(X^*)] - \frac{\mu - m}{s^2} \mathbf{E}[\psi_A(X)] \\ &= \int_{S(X^*)} \left[f(x) \frac{h^*(x)}{h(x)} - \frac{\sigma^2}{s^2} f^*(x) \right] \psi'_A(x) dx - \frac{\mu - m}{s^2} \mathbf{E}[\psi_A(X)]. \end{aligned}$$

Therefore, since $S(X) \subset S(X^*) \subset S(Y) = (a, b)$, we have from Lemma 2.1 that

$$\begin{aligned} |\mathbf{P}\{X \in A\} - \mathbf{P}\{Y \in A\}| &\leq \operatorname{ess\,sup}_{x \in (a, b)} \left| \frac{h^*(x)}{h(x)} \psi'_A(x) \right| \int_a^b \left| f(x) - \frac{\sigma^2 h(x)}{s^2 h^*(x)} f^*(x) \right| dx \\ &+ \frac{|\mu - m|}{s^2} \sup_{x \in (a, b)} |\psi_A(x)| \leq 2 \int_a^b \left| f(x) - \frac{\sigma^2 h(x)}{s^2 h^*(x)} f^*(x) \right| dx + c_Y |\mu - m|, \end{aligned}$$

which is independent of A , and the result follows.

In particular, for the normal r.v. $Z_{m, \sigma^2} = \sigma Z + m$, where Z is standard normal, the following result holds.

Corollary 2.1. *For any a.c. r.v. X with mean μ and variance σ^2 ,*

$$d_{TV}(X, Z_{m, \sigma^2}) \leq 3 d_{TV}(X, X^*) + \frac{\sqrt{\pi}}{\sigma \sqrt{2}} |\mu - m|. \quad (2.5)$$

Remark 2.1. It should be noted that (2.4) immediately yields (2.5) with a constant 4 instead of 3.

Proof of Corollary 2.1. From (2.3) we have that for all x and A ,

$$|(x - m) \psi_A(x)| \leq \sigma^2.$$

On the other hand, the normal density $h(x) = \varphi(x; \mu, \sigma^2) = (1/\sigma) \varphi((x - m)/\sigma)$, where φ is the standard normal density, satisfies $h^* \equiv h$. Therefore, from the calculations in the proof of Theorem 2.1, it follows that

$$\begin{aligned} \mathbf{P}\{X \in A\} - \mathbf{P}\{Z_{m, \sigma^2} \in A\} &= \int_{S(X^*)} (f(x) - f^*(x)) \psi'_A(x) dx - \frac{\mu - m}{\sigma^2} \mathbf{E}[\psi_A(X)] \\ &= \int_{S(X^*)} \frac{x - m}{\sigma^2} \psi_A(x) (f(x) - f^*(x)) dx \\ &\quad + \int_{S(X^*)} I(x \in A) (f(x) - f^*(x)) dx - \frac{\mu - m}{\sigma^2} \mathbf{E}[\psi_A(X)]. \end{aligned}$$

Observe that for any Borel set A ,

$$\int_{S(X^*)} I(x \in A) (f(x) - f^*(x)) dx \leq d_{TV}(X, X^*)$$

and

$$\int_{S(X^*)} \frac{x - m}{\sigma^2} \psi_A(x) (f(x) - f^*(x)) dx \leq 2 d_{TV}(X, X^*).$$

The result follows from the fact that $\mathbf{E}|Z_{m, \sigma^2} - m| = \sigma \sqrt{2/\pi}$ and the observation that for any two r.v.'s X_1 and X_2 ,

$$d_{TV}(X_1, X_2) = \sup_A \left\{ \mathbf{P}\{X_1 \in A\} - \mathbf{P}\{X_2 \in A\}, A \text{ Borel} \right\}.$$

As a consequence, a necessary and sufficient condition for an a.c. r.v. X with a finite second moment to have a normal distribution is that $X \stackrel{d}{=} X^*$ (see also [9]). The stability of this result is shown in the following theorem.

Theorem 2.2. *Let X_n be a sequence of a.c. r.v.'s with means μ_n and variances σ_n^2 such that $\mu_n \rightarrow \mu$ and $\sigma_n^2 \rightarrow \sigma^2$ as $n \rightarrow \infty$, where $0 < \sigma^2 < \infty$. Then, as $n \rightarrow \infty$,*

(i) $d_{TV}(X_n, Z_{\mu, \sigma^2}) \rightarrow 0$ if and only if $d_{TV}(X_n, X_n^*) \rightarrow 0$.

(ii) $d_{TV}(X_n, Z_{\mu, \sigma^2}) \rightarrow 0$ implies $d_{TV}(X_n^*, Z_{\mu, \sigma^2}) \rightarrow 0$.

Proof. Assume first that $d_{TV}(X_n, X_n^*) \rightarrow 0$. Then, we simply have from Corollary 2.1 and Theorem 2.1 that

$$\begin{aligned} d_{TV}(X_n, Z_{\mu, \sigma^2}) &\leq d_{TV}(X_n, Z_{\mu, \sigma_n^2}) + d_{TV}(Z_{\mu, \sigma_n^2}, Z_{\mu, \sigma^2}) \\ &\leq 3 d_{TV}(X_n, X_n^*) + \frac{\sqrt{\pi}}{\sigma_n \sqrt{2}} |\mu_n - \mu| + 2 \left| 1 - \frac{\sigma_n^2}{\sigma^2} \right| \rightarrow 0. \end{aligned}$$

Assume now that $d_{TV}(X_n, Z_{\mu, \sigma^2}) \rightarrow 0$. It follows that every subsequence $\{m\} \subset \{n\}$ contains a further subsequence $\{k\} \subset \{m\}$ such that for almost all x , $f_k(x) \rightarrow \varphi(x; \mu, \sigma^2)$ as $k \rightarrow \infty$, where f_k is the density of X_k . Therefore, for almost all x ,

$$\frac{(\mu_k - x)^2}{\sigma_k^2} f_k(x) \rightarrow \frac{(\mu - x)^2}{\sigma^2} \varphi(x; \mu, \sigma^2) \quad \text{as } k \rightarrow \infty,$$

and from Scheffé's lemma,

$$\int_{-\infty}^{\infty} \left| \frac{(\mu_k - x)^2}{\sigma_k^2} f_k(x) - \frac{(\mu - x)^2}{\sigma^2} \varphi(x; \mu, \sigma^2) \right| dx \rightarrow 0. \quad (2.6)$$

By using Tonelli's theorem, we have

$$\begin{aligned} 2d_{TV}(X_k^*, Z_{\mu_k, \sigma_k^2}) &= \int_{-\infty}^{\mu_k} \left| \int_{-\infty}^x \frac{\mu_k - t}{\sigma_k^2} (f_k(t) - \varphi(t; \mu_k, \sigma_k^2)) dt \right| dx \\ &\quad + \int_{\mu_k}^{\infty} \left| \int_x^{\infty} \frac{t - \mu_k}{\sigma_k^2} (f_k(t) - \varphi(t; \mu_k, \sigma_k^2)) dt \right| dx \\ &\leq \int_{-\infty}^{\mu_k} \int_{-\infty}^x \frac{\mu_k - t}{\sigma_k^2} |f_k(t) - \varphi(t; \mu_k, \sigma_k^2)| dt dx \\ &\quad + \int_{\mu_k}^{\infty} \int_x^{\infty} \frac{t - \mu_k}{\sigma_k^2} |f_k(t) - \varphi(t; \mu_k, \sigma_k^2)| dt dx \\ &= \int_{-\infty}^{\infty} \frac{(\mu_k - t)^2}{\sigma_k^2} |f_k(t) - \varphi(t; \mu_k, \sigma_k^2)| dt \\ &\leq \int_{-\infty}^{\infty} \left| \frac{(\mu_k - t)^2}{\sigma_k^2} f_k(t) - \frac{(\mu - t)^2}{\sigma^2} \varphi(t; \mu, \sigma^2) \right| dt \\ &\quad + \int_{-\infty}^{\infty} \left| \frac{(\mu - t)^2}{\sigma^2} \varphi(t; \mu, \sigma^2) - \frac{(\mu_k - t)^2}{\sigma_k^2} \varphi(t; \mu_k, \sigma_k^2) \right| dt \rightarrow 0, \end{aligned}$$

as $k \rightarrow \infty$, by (2.6). Therefore,

$$d_{TV}(X_k, X_k^*) \leq d_{TV}(X_k, Z_{\mu, \sigma^2}) + d_{TV}(Z_{\mu, \sigma^2}, Z_{\mu_k, \sigma_k^2}) + d_{TV}(X_k^*, Z_{\mu_k, \sigma_k^2}) \rightarrow 0,$$

as $k \rightarrow \infty$, and the «only if» part is complete. Regarding (ii), we simply have from (i)

$$d_{TV}(X_n^*, Z_{\mu, \sigma^2}) \leq d_{TV}(X_n, X_n^*) + d_{TV}(X_n, Z_{\mu, \sigma^2}) \rightarrow 0,$$

which completes the proof.

Remark 2.2. It should be noted that (2.4) extends the results of [14]. Indeed, if X has an interval support, (2.4) yields

$$d_{TV}(X, Y) \leq 2\mathbf{E} \left| 1 - \frac{\sigma^2 w(X)}{s^2 w_Y(X)} \right| + c_Y |\mu - m|,$$

where $w = f^*/f$ and $w_Y = h^*/h$ are the w -functions of X and Y , respectively.

3. An application to the local limit theorem for densities. Let $X, X_1, X_2, \dots$ be a sequence of i.i.d. r.v.'s with mean zero, variance one, and a common density f , and consider the partial sums $S_n = X_1 + \dots + X_n$ with densities f_n . In the present section we give a simple proof of the following result.

Theorem 3.1. *If Z is the standard normal r.v., then*

$$d_{TV}\left(\frac{S_n}{\sqrt{n}}, Z\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (3.1)$$

It should be noted that several proofs of Prokhorov's theorem (3.1), have appeared in the probability literature the last few years (see for example [1], [13], [6]), but it seems that the authors use some additional restrictive conditions on the density f . The present approach, however, does not require any further assumption, other than a finite second moment. Our approach extends the results of [6], dispensing with the assumption of an interval support. The proof will make use of the following lemmas.

Lemma 3.1. We have $S_n^* \stackrel{d}{=} S_{n-1}^* + X_n$, where S_{n-1}^* is independent of X_n .

P r o o f. From Lemma 1.1(v) we have $(X_1 + X_2)^* \stackrel{d}{=} X_1^* + X_2$ which proves the lemma for $n = 2$. By induction on n ,

$$S_n^* \stackrel{d}{=} X_1^* + (X_2 + \cdots + X_n). \quad (3.2)$$

Therefore,

$$S_{n-1}^* + X_n \stackrel{d}{=} X_1^* + (X_2 + \cdots + X_{n-1}) + X_n \stackrel{d}{=} S_n^*.$$

Lemma 3.2. The sequence $d_{TV}(S_n, S_n^*)$ is nonincreasing and, hence, it converges to a nonnegative constant d .

P r o o f. We simply have from Lemma 3.1 that

$$d_{TV}(S_n, S_n^*) = d_{TV}(S_{n-1} + X_n, S_{n-1}^* + X_n) \leq d_{TV}(S_{n-1}, S_{n-1}^*).$$

Lemma 3.3. If $S(X) = \{x: f(x) > 0\}$ is the support of X , then

$$d \leq 1 - \mathbf{P}\{X^* \in S(X)\}.$$

P r o o f. Let G_n , $n = 1, 2, \dots$, be a sequence of arbitrary measurable functions such that $0 \leq G_n(x) \leq 1$ for all n and x . Using (3.2), we have

$$\begin{aligned} \mathbf{E}[G_n(S_n^*)] &= \mathbf{E}[G_n(X_1^* + (X_2 + \cdots + X_n))] \\ &\geq \mathbf{E}[G_n(X_1^* + (X_2 + \cdots + X_n)) I(X_1^* \in S(X))] \\ &= \mathbf{E}\left[\frac{f^*(X_1)}{f(X_1)} G_n(S_n)\right] = \frac{1}{n} \sum_{j=1}^n \mathbf{E}\left[\frac{f^*(X_j)}{f(X_j)} G_n(S_n)\right], \end{aligned}$$

where f^* is the (common) density of X_j^* . Therefore,

$$\begin{aligned} \mathbf{E}[G_n(S_n)] - \mathbf{E}[G_n(S_n^*)] &\leq \mathbf{E}\left[\left(1 - \frac{1}{n} \sum_{j=1}^n \frac{f^*(X_j)}{f(X_j)}\right) G_n(S_n)\right] \\ &= \mathbf{E}\left[\left(\mathbf{P}\{X^* \in S(X)\} - \frac{1}{n} \sum_{j=1}^n \frac{f^*(X_j)}{f(X_j)}\right) G_n(S_n)\right] \\ &\quad + (1 - \mathbf{P}\{X^* \in S(X)\}) \mathbf{E}[G_n(S_n)] \\ &\leq \mathbf{E}\left|\mathbf{P}\{X^* \in S(X)\} - \frac{1}{n} \sum_{j=1}^n \frac{f^*(X_j)}{f(X_j)}\right| + 1 - \mathbf{P}\{X^* \in S(X)\}. \end{aligned}$$

Since it is easily verified that the nonnegative i.i.d. r.v.'s $Y_j = f^*(X_j)/f(X_j)$ have the mean $\mathbf{E}[Y_j] = \mathbf{P}\{X^* \in S(X)\}$, it follows from the weak law of large numbers that

$$\frac{1}{n} \sum_{j=1}^n Y_j \xrightarrow{P} \mathbf{P}\{X^* \in S(X)\} = \mathbf{E}\left[\frac{1}{n} \sum_{j=1}^n Y_j\right] \quad \text{as } n \rightarrow \infty.$$

Therefore, we conclude (see for example [8, p. 101, Corollary 4]) that

$$\mathbf{E}\left|\mathbf{P}\{X^* \in S(X)\} - \frac{1}{n} \sum_{j=1}^n Y_j\right| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

The proof is completed by choosing $G_n(x) = I(x \in A_n)$, where f_n^* is the density of S_n^* and $A_n = \{x: f_n(x) > f_n^*(x)\}$.

The following lemma (which is probably known) will be used in the sequel. Since, however, we have not been able to trace a proof, we provide one for the completeness of the presentation.

Lemma 3.4. Let Y, W be independent and a.c. r.v.'s with densities h, g , respectively, and let $q = h * g$ be the density of $Y + W$. Then we have the following:

- (i) The set $S(Y + W) = \{x: q(x) > 0\}$ contains some interval;
- (ii) If $0 < \mathbf{P}\{Y \leq 0\} < 1$ and $0 < \mathbf{P}\{W \leq 0\} < 1$, then there exist four positive numbers $a_1 < a_2$ and $b_1 < b_2$ such that $(-a_2, -a_1) \cup (b_1, b_2) \subset S(Y + W)$;
- (iii) If $(a, b) \subset S(Y) = \{x: h(x) > 0\}$ and $(c, d) \subset S(W) = \{x: g(x) > 0\}$, then $(a, b) + (c, d) = (a + c, b + d) \subset S(Y + W)$.

P r o o f. Set $h_k(x) = \min\{k, h(x)\}$ for $k = 1, 2, \dots$. Obviously h_k is bounded and thus $h_k * g$ is continuous. Therefore, the set $A_k = \{x: (h_k * g)(x) > 0\}$ is open. Since $h_k \uparrow h$ as $k \rightarrow \infty$ and $h_k * g \leq q$ for all k , it follows that $A_k \subset S(Y + W)$. By monotone convergence, $h_k * g \uparrow q$ as $k \rightarrow \infty$, which shows that for large enough k , $(h_k * g)(x) > 0$ if $q(x) > 0$. Therefore, A_k is nonempty for large enough k and hence it contains an interval; this proves (i). As regards (ii), observe that the set $S(Y + W)$ contains positive and negative numbers (otherwise $\mathbf{P}\{Y + W \leq 0\}$ would be equal to 0 or 1), and the result follows as in (i). Finally (iii) is trivial.

It is easy to see that for any two r.v.'s as in Lemma 3.4, $S(Y + W) \subset S(Y) + S(W)$, where $S(Y) + S(W) = \{y + w: y \in S(Y), w \in S(W)\}$. Therefore, applying Lemma 3.4(i) to the independent r.v.'s Y, W with densities $h(x) = I(x \in A)/\lambda(A)$ and $g(x) = I(x \in B)/\lambda(B)$ (where $0 < \lambda(A) < \infty$ and $0 < \lambda(B) < \infty$) we conclude the classical result, due to Steinhaus (see for example [5]), that $S(Y) + S(W) = A + B$ contains an interval.

As an immediate consequence of Lemma 3.4, the following result holds.

Corollary 3.1. There exists an integer m and a positive a such that $(-na, na) \subset S(S_{nm})$ for $n = 1, 2, \dots$, where $S(S_{nm})$ is the support of S_{nm} .

P r o o f. From Lemma 3.4(ii) we have $(-a_2, -a_1) \cup (b_1, b_2) \subset S(X_1 + X_2) = S(S_2)$. By using Lemma 3.4(iii) and the fact that $S_{2n} = (X_1 + X_2) + \dots + (X_{2n-1} + X_{2n})$, we conclude that

$$\bigcup_{k=0}^n (-ka_2 + (n-k)b_1, -ka_1 + (n-k)b_2) \subset S(S_{2n}).$$

Choose now two integers $k < m'$ such that $a_1/b_2 + 1 < m'/k < a_2/b_1 + 1$. Then, there exists a positive a such that $(-a, a) \subset S(S_{2m'})$ and thus, $(-na, na) \subset S(S_{nm})$ for all $n \geq 1$, where $m = 2m'$.

Finally, we will use the following lemma (for a proof see [10, p. 19] or [2, p. 176]).

Lemma 3.5. Let $R, R_1, R_2, \dots$ be i.i.d. r.v.'s with mean zero and finite variance $\tau^2 > 0$. Then, for any $a > 0$,

$$\frac{1}{n} \mathbf{E} [T_n^2 I(|T_n| > na)] \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

where $T_n = R_1 + \dots + R_n$.

Although the proof of Lemma 3.5 is far from obvious, it becomes simpler when $\mathbf{E}[|R|^\delta] < \infty$ for some $\delta > 2$ (it follows by an application of Hölder's inequality with $p = \delta/2$ and a further application of Marcinkiewicz–Zygmund inequality, [10, p. 169], (2.2) and (2.3)). We can now prove the main result.

P r o o f o f T h e o r e m 3.1. Let a and m be as in Corollary 3.1. It follows that

$$\begin{aligned} \mathbf{P}\{S_{nm}^* \in S(S_{nm})\} &= \int_{S(S_{nm})} f_{nm}^*(x) dx \geq \int_{-na}^{na} f_{nm}^*(x) dx \\ &= \frac{1}{nm} \left\{ na \int_{|x| > na} |x| f_{nm}(x) dx + \int_{-na}^{na} x^2 f_{nm}(x) dx \right\} \\ &= \frac{a}{m} \mathbf{E}[|S_{nm}| I(|S_{nm}| > na)] + \frac{1}{nm} \mathbf{E}[S_{nm}^2 I(|S_{nm}| \leq na)] \\ &= 1 + \frac{a}{m} \mathbf{E}[|S_{nm}| I(|S_{nm}| > na)] - \frac{1}{nm} \mathbf{E}[S_{nm}^2 I(|S_{nm}| > na)] \\ &\geq 1 - \frac{1}{nm} \mathbf{E}[S_{nm}^2 I(|S_{nm}| > na)] \rightarrow 1 \quad \text{as } n \rightarrow \infty, \end{aligned}$$

by Lemma 3.5. Let R_j be i.i.d. r.v.'s such that $R_j \stackrel{d}{=} S_{mk}/\sqrt{mk}$, $j = 1, 2, \dots$. From Lemmas 3.3 and 3.2 it follows that

$$\begin{aligned} d &= \lim_{n \rightarrow \infty} d_{TV}(S_{nmk}, S_{nmk}^*) = \lim_{n \rightarrow \infty} d_{TV}\left(\frac{S_{nmk}}{\sqrt{mk}}, \frac{S_{nmk}^*}{\sqrt{mk}}\right) \\ &\leq 1 - \mathbf{P}\{R_1^* \in S(R_1)\} = 1 - \mathbf{P}\{S_{mk}^* \in S(S_{mk})\} \rightarrow 0, \quad \text{as } k \rightarrow \infty, \end{aligned}$$

which shows that $d = 0$. The result follows from Corollary 2.1, since

$$d_{TV}\left(\frac{S_n}{\sqrt{n}}, Z\right) \leq 3 d_{TV}\left(\frac{S_n}{\sqrt{n}}, \frac{S_n^*}{\sqrt{n}}\right) = 3 d_{TV}(S_n, S_n^*) \rightarrow 0.$$

REFERENCES

1. Barron A. R. Entropy and the central limit theorem. — Ann. Probab., 1986, v. 14, p. 336–342.
2. Billingsley P. Convergence of Probability Measures, New York: Wiley, 1968.
3. Bolthausen E. An estimate of the remainder in a combinatorial central limit theorem. — Z. Wahrscheinlichkeitstheor. verw. Geb., 1984, B. 66, S. 379–386.
4. Боровков А. А., Утев С. А. Об одном неравенстве и связанной с ним характеристикой нормального распределения. — Теория вероятн. и ее примен., 1983, т. 28, в. 2, с. 209–218.
5. Brown G. Basic measures. — Proc. Centre for Mathem. Analysis, Australian National University, McIntosh, A. G. R. and Pryde A. J., Eds, 1985, v. 9, p. 132–140.
6. Cacoullos T., Papathanasiou V., Utev S. Variational inequalities with examples and an application to the central limit theorem. — Ann. Probab., 1994, v. 22, p. 1607–1618.
7. Chen L. H. Y. The central limit theorem and Poincaré type inequalities. — Ann. Probab., 1988, v. 16, p. 300–304.
8. Chow Y. S., Teicher H. Probability Theory: Independence, Interchangeability, Martingales. New York: Springer-Verlag, 2nd ed., 1988.
9. Goldstein L., Reinert R. Stein's method and the zero bias transformation with application to simple random sampling. — Ann. Appl. Probab., 1997, v. 7, № 4, p. 935–952.
10. Gut A. Stopped random walks: Limit Theorems and Applications. New York: Springer-Verlag, 1988.
11. Hudson H. M. A natural identity for exponential families with applications to multi-parameter estimation. — Ann. Statist., 1978, v. 6, p. 473–484.
12. Liu J. S. Siegel's formula via Stein's identities. — Statist. Probab. Lett., 1994, v. 21, p. 247–251.
13. Mayer-Wolf E. The Cramer-Rao functional and limiting laws. — Ann. Probab., 1990, v. 18, № 2, p. 840–850.
14. Papadatos N., Papathanasiou V. Distance in variation between two arbitrary distributions via the associated w -functions. — Теория вероятн. и ее примен., 1995, т. 40, в. 3, с. 685–694.
15. Papadatos N., Papathanasiou V. Unified variance bounds and a Stein-type identity. — Technical Report TR/17/98, Dept. of Mathematics and Statistics, University of Cyprus, 1998.
16. Прохоров Ю. В. О локальной предельной теореме для плотностей. — Докл. АН СССР, 1952, т. 83, с. 797–800.
17. Stein C. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. — Proc. Sixth Berkeley Symp. Math. Statist. Probab., Berkeley, Univ. California Press, 1972, v. 2, p. 583–602.
18. Stein C. Estimation of the mean of a multivariate normal distribution. — Ann. Statist., 1981, v. 9, p. 1135–1151.