THEORY PROBAB. APPL.
Vol. 46, No. 4

AN APPLICATION OF A DENSITY TRANSFORM AND THE LOCAL LIMIT THEOREM*

T. CACOULLOS†, N. PAPADATOS‡, AND V. PAPATHANASIOU‡§

Abstract. Consider an absolutely continuous random variable X with finite variance σ^2. It is known that there exists another random variable X^* (which can be viewed as a transformation of X) with a unimodal density, satisfying the extended Stein-type covariance identity $\mathrm{Cov}[X,g(X)] = \sigma^2\mathbf{E}[g'(X^*)]$ for any absolutely continuous function g with derivative g', provided that $\mathbf{E}|g'(X^*)| < \infty$. Using this transformation, upper bounds for the total variation distance between two absolutely continuous random variables X and Y are obtained. Finally, as an application, a proof of the local limit theorem for sums of independent identically distributed random variables is derived in its full generality.

Key words. density transform, total variation distance, local limit theorem for densities

PII. S0040585X97979366

1. Introduction.

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

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

The following known properties of X^* (see [11] or [17]) will be used in what follows.

LEMMA 1.1. (i) *The density f^* is unimodal and absolutely continuous with mode μ and maximal value*

$$f^*(\mu) = \frac{\mathbf{E}|X-\mu|}{2\sigma^2}.$$

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

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

Moreover, X^* is characterized by this property.

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

(iv) *For any real numbers $a \neq 0$ and b, $(aX+b)^* \stackrel{\mathrm{d}}{=} aX^* + b$* (throughout this paper, $X \stackrel{\mathrm{d}}{=} Y$ 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 $a_1a_2 \neq 0$,*

$$(a_1X_1 + a_2X_2)^* \stackrel{\mathrm{d}}{=} B(a_1X_1^* + a_2X_2) + (1-B)(a_1X_1 + a_2X_2^*), \tag{1.3}$$

where X_1, X_2, X_1^*, X_2^*, and B are mutually independent with

$$\mathbf{P}\{B=1\} = \frac{a_1^2\sigma_1^2}{a_1^2\sigma_1^2 + a_2^2\sigma_2^2} = 1 - \mathbf{P}\{B=0\}.$$

*Received by the editors January 23, 1999. This work was supported by the Ministry of Industry, Energy and Technology of Greece, grant 1369.

http://www.siam.org/journals/tvp/46-4/97936.html

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

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

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

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 [22], [21]). The classical Stein identity has had many interesting applications to several areas of probability and statistics (see, e.g., [13] and [14]). Other important applications include the well-known Stein method for the approximation of the distribution of a sum of dependent r.v.'s [21]. Another point of view is given by Chen [8], where the CLT is obtained by using Poincaré-type inequalities (closely connected with this identity), and by the results of [4].

Some applications of (1.2), concerning variance bounds and characterizations of distributions, are discussed in [17]. Moreover, Goldstein and Reinert [11] extended Stein's method and 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 [11, Corollary 3.1]). Furthermore, they used the so-called *zero biased transformation* of X (i.e., the r.v. X^*) to obtain estimates for the rate of convergence for several dependent samples.

The purpose of this paper is to obtain 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 [16] are extended in two directions: (i) we do not require 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 [11, 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 [20], for sums of i.i.d. r.v.'s, is given in section 3.

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

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

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" r.v. 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 [16]) 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 $-\infty \leqq a < b \leqq \infty$. Fix an arbitrary Borel set A and consider the a.c. function (c.f. [3], [6], [16])

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

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*

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

Proof. Let $a < x < x + y \leqq m$. It follows that

$$\int_a^x h(t)\,dt \int_x^{x+y} (m-t)\,h(t)\,dt \leqq (m-x) \int_a^x h(t)\,dt \int_x^{x+y} h(t)\,dt$$
$$\leqq \int_a^x (m-t)\,h(t)\,dt \int_x^{x+y} h(t)\,dt.$$

By adding to both sides the quantity $\int_a^x h(t)\,dt \int_a^x (m-t)\,h(t)\,dt$ we conclude that

$$s^2 h^*(x)\,H(x+y) \geqq s^2 h^*(x+y)\,H(x),$$

where H is the distribution function of Y. Since $x > a$, $h^*(x) > 0$, and $H(x) > 0$, it follows that the function $H(x)/h^*(x)$ is nondecreasing for $x \in (a, m]$. Similarly, $(1 - H(x))/h^*(x)$ is nonincreasing for $x \in [m, b)$. Since

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

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

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

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

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

Hence,

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

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

$$\left|\frac{h^*(x)}{h(x)}\,\psi_A'(x)\right| \leqq \frac{(m-x)\int_a^x h(t)\,dt}{\int_a^x (m-t)\,h(t)\,dt} + 1 \leqq 2. \tag{2.3}$$

Similarly, $|h^*(x)\psi_A'(x)/h(x)| \leqq 2$ 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,*

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

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

Proof. 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. [16])

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

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

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

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,*

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

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,

$$\bigl|(x - m) \, \psi_A(x)\bigr| \leqq \sigma^2.$$

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

$$\mathbf{P}\{X \in A\} - \mathbf{P}\{Z_{m,\sigma^2} \in A\} = \int_{S(X^*)} \bigl(f(x) - f^*(x)\bigr) \, \psi_A'(x) \, dx - \frac{\mu - m}{\sigma^2} \, \mathbf{E}\bigl[\psi_A(X)\bigr]$$
$$= \int_{S(X^*)} \frac{x - m}{\sigma^2} \, \psi_A(x) \bigl(f(x) - f^*(x)\bigr) \, dx$$
$$+ \int_{S(X^*)} I(x \in A) \, \bigl(f(x) - f^*(x)\bigr) \, dx - \frac{\mu - m}{\sigma^2} \, \mathbf{E}\bigl[\psi_A(X)\bigr].$$

Observe that for any Borel set A,

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

and

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

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,

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

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 $X \overset{\mathrm{d}}{=} X^*$ (see also [11]). 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 $\mu_n \to \mu$ and $\sigma_n^2 \to \sigma^2$ as $n \to \infty$, where $0 < \sigma^2 < \infty$. Then, as $n \to \infty$,*

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

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

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

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

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

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

and from Scheffé's lemma,

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

By using Tonelli's theorem, we have

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

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

$$d_{TV}\left(X_k, X_k^*\right) \leqq d_{TV}\left(X_k, Z_{\mu,\sigma^2}\right) + d_{TV}\left(Z_{\mu,\sigma^2}, Z_{\mu_k,\sigma_k^2}\right) + d_{TV}\left(X_k^*, Z_{\mu_k,\sigma_k^2}\right) \longrightarrow 0,$$

as $k \to \infty$, and the "only if" part is complete. Regarding (ii), we simply have from (i) that

$$d_{TV}\left(X_n^*, Z_{\mu,\sigma^2}\right) \leqq d_{TV}\left(X_n, X_n^*\right) + d_{TV}\left(X_n, Z_{\mu,\sigma^2}\right) \longrightarrow 0,$$

which completes the proof.

Remark 2.2. It should be noted that (2.4) extends the results of [16]. Indeed, if X has an interval support, (2.4) yields

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

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 $X, X_1, X_2, \ldots$ 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 $S_n = X_1 + \cdots + X_n$ 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*

$$d_{TV}\left(\frac{S_n}{\sqrt{n}},\ Z\right) \longrightarrow 0 \qquad as \quad n \to \infty. \tag{3.1}$$

It should be noted that several proofs of Prokhorov's theorem (3.1) have appeared in the probability literature within the last 20 years (see, for example, [1], [15], [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 $S_n^* \stackrel{\mathrm{d}}{=} S_{n-1}^* + X_n$, where S_{n-1}^* is independent of X_n.*

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

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

Therefore,

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

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

Proof. We simply have from Lemma 3.1 that

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

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

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

Proof. Let G_n, $n = 1, 2, \ldots$, be a sequence of arbitrary measurable functions such that $0 \leqq G_n(x) \leqq 1$ for all n and x. Using (3.2), we have

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

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

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

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

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

Therefore, we conclude (see, for example, [10, Corollary 4, p. 101]) that

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

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

The next lemma (which is probably known) will be used in what follows. 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 $S(Y+W) = \{x\colon\ q(x) > 0\}$ contains some interval.*

(ii) *If $0 < \mathbf{P}\{Y \leqq 0\} < 1$ and $0 < \mathbf{P}\{W \leqq 0\} < 1$, 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 $(a,b) \subset S(Y) = \{x\colon\ h(x) > 0\}$ and $(c,d) \subset S(W) = \{x\colon\ g(x) > 0\}$, then $(a,b) + (c,d) = (a+c, b+d) \subset S(Y+W)$.*

Proof. Set $h_k(x) = \min\{k, h(x)\}$ for $k = 1, 2, \ldots$. Obviously h_k is bounded and thus $h_k * g$ is continuous. Therefore, the set $A_k = \{x\colon\ (h_k * g)(x) > 0\}$ is open. Since $h_k \uparrow h$ as $k \to \infty$ and $h_k * g \leqq q$ for all k, it follows that $A_k \subset S(Y+W)$. By monotone convergence, $h_k * g \uparrow q$ as $k \to \infty$, 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 contains an interval; this proves (i). Regarding (ii), observe that the set $S(Y+W)$ contains positive and negative numbers (otherwise $\mathbf{P}\{Y+W \leqq 0\}$ 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 $S(Y) + S(W) = \{y + w\colon\ y \in S(Y), w \in S(W)\}$. 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 $n = 1, 2, \ldots$, where $S(S_{nm})$ is the support of S_{nm}.*

Proof. 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 $S_{2n} = (X_1 + X_2) + \cdots + (X_{2n-1} + X_{2n})$, we conclude that

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

Now choose 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 $n \geqq 1$, where $m = 2m'$.

Finally, we will use the following lemma (for a proof see [12, p. 19] or [2, p. 176]).

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

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

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

Although the proof of Lemma 3.5 is not 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 the Marcinkiewicz–Zygmund inequality [12, p. 169], (2.2), and (2.3)). We can now prove the main result.

Proof of Theorem 3.1. Let a and m be as in Corollary 3.1. It follows that by Lemma 3.5

$$\begin{aligned}\mathbf{P}\big\{S^*_{nm} \in S(S_{nm})\big\} &= \int_{S(S_{nm})} f^*_{nm}(x)\,dx \geqq \int_{-na}^{na} f^*_{nm}(x)\,dx \\ &= \frac{1}{nm}\left\{na\int_{|x|>na} |x|\, f_{nm}(x)\,dx + \int_{-na}^{na} x^2\, f_{nm}(x)\,dx\right\} \\ &= \frac{a}{m}\,\mathbf{E}\big[|S_{nm}|\, I\big(|S_{nm}| > na\big)\big] + \frac{1}{nm}\,\mathbf{E}\big[S^2_{nm} I\big(|S_{nm}| \leqq na\big)\big] \\ &= 1 + \frac{a}{m}\,\mathbf{E}\big[|S_{nm}|\, I\big(|S_{nm}| > na\big)\big] - \frac{1}{nm}\,\mathbf{E}\big[S^2_{nm} I\big(|S_{nm}| > na\big)\big] \\ &\geqq 1 - \frac{1}{nm}\,\mathbf{E}\big[S^2_{nm} I\big(|S_{nm}| > na\big)\big] \longrightarrow 1 \qquad \text{as} \quad n \to \infty.\end{aligned}$$

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

$$\begin{aligned}d = \lim_{n\to\infty} d_{TV}(S_{nmk}, S^*_{nmk}) &= \lim_{n\to\infty} d_{TV}\left(\frac{S_{nmk}}{\sqrt{mk}},\ \frac{S^*_{nmk}}{\sqrt{mk}}\right) \\ \leqq 1 - \mathbf{P}\big\{R^*_1 \in S(R_1)\big\} &= 1 - \mathbf{P}\big\{S^*_{mk} \in S(S_{mk})\big\} \longrightarrow 0 \qquad \text{as} \quad k \to \infty,\end{aligned}$$

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

$$d_{TV}\left(\frac{S_n}{\sqrt{n}},\ Z\right) \leqq 3\, d_{TV}\left(\frac{S_n}{\sqrt{n}},\ \frac{S^*_n}{\sqrt{n}}\right) = 3\, d_{TV}\left(S_n,\ S^*_n\right) \longrightarrow 0.$$

REFERENCES

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

[2] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[3] E. Bolthausen, *An estimate of the remainder in a combinatorial central limit theorem*, Z. Wahrsch. Verw. Gebiete, 66 (1984), pp. 379–386.

[4] A. A. Borovkov and S. A. Utev, *On an inequality and a related characterization of the normal distribution*, Theory Probab. Appl., 28 (1983), pp. 219–228.

[5] G. Brown *Basic measures*, in Miniconference on Linear Analysis and Functional Spaces (Canberra, 1984), Proc. Centre Math. Anal. Austral. Nat. Univ. 9, Australian National University, Canberra, 1985, pp. 132–140.

[6] T. Cacoullos, V. Papathanasiou, and S. Utev, *Variational inequalities with examples and an application to the central limit theorem*, Ann. Probab., 22 (1994), pp. 1607–1618.

[7] L. H. Y. Chen, *An inequality for the multivariate normal distribution*, J. Multivariate Anal., 12 (1982), pp. 306–315.

[8] L. H. Y. Chen, *The central limit theorem and Poincaré type inequalities*, Ann. Probab., 16 (1988), pp. 300–304.

[9] H. Chernoff, *A note on an inequality involving the normal distribution*, Ann. Probab., 9 (1981), pp. 533–535.

[10] Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, 2nd ed., Springer-Verlag, New York, 1988.

[11] L. Goldstein and R. Reinert, *Stein's method and the zero bias transformation with application to simple random sampling*, Ann. Appl. Probab., 7 (1997), pp. 935–952.

[12] A. Gut, *Stopped Random Walks: Limit Theorems and Applications*, Springer-Verlag, New York, 1988.

[13] H. M. HUDSON, *A natural identity for exponential families with applications to multiparameter estimation*, Ann. Statist., 6 (1978), pp. 473–484.

[14] J. S. LIU, *Siegel's formula via Stein's identities*, Statist. Probab. Lett., 21 (1994), pp. 247–251.

[15] E. MAYER-WOLF, *The Cramer-Rao functional and limiting laws*, Ann. Probab., 18 (1990), pp. 840–850.

[16] N. PAPADATOS AND V. PAPATHANASIOU, *Distance in variation between two arbitrary distributions via the associated w-functions*, Theory Probab. Appl., 40 (1995), pp. 567–575.

[17] N. PAPADATOS AND V. PAPATHANASIOU, *Unified Variance Bounds and a Stein-type Identity*, Technical Report TR/17/98, Dept. of Mathematics and Statistics, University of Cyprus, 1998.

[18] V. PAPATHANASIOU, *Characterizations of multidimensional exponential families by Cacoullos-type inequalities*, J. Multivariate Anal., 35 (1990), pp. 102–107.

[19] V. PAPATHANASIOU, *Some characteristic properties of the Fisher information matrix via Cacoullos-type inequalities*, J. Multivariate Anal., 44 (1993), pp. 256–265.

[20] YU. V. PROKHOROV, *On a local limit theorem for densities*, Dokl. Acad. Nauk. USSR, 83 (1952), pp. 797–800 (in Russian).

[21] C. STEIN, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, in Proc. Sixth Berkeley Symp. Math. Statist. Probab. (Berkeley, CA, 1970/1971), University California Press, Berkeley, CA, 1972, pp. 583–602.

[22] C. STEIN, *Estimation of the mean of a multivariate normal distribution*, Ann. Statist., 9 (1981), pp. 1135–1151.