PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 95, Number 4, December 1985

A TOOL IN ESTABLISHING TOTAL VARIATION CONVERGENCE[1]

K. R. PARTHASARATHY AND TON STEERNEMAN

ABSTRACT. Let $X_0, X_1, X_2, \ldots$ and $Y_0, Y_1, Y_2, \ldots$ be sequences of random variables where X_n and Y_n are independent, $\mathscr{L}X_n \to \mathscr{L}X_0$ in total variation and $\mathscr{L}Y_n \to \mathscr{L}Y_0$ in distribution. For certain mappings T sufficient conditions are given in order that $\mathscr{L}T(X_n, Y_n) \to \mathscr{L}T(X_0, Y_0)$ in total variation. For example, if $(\mathbf{R}^k, \mathscr{B}_k)$ is the outcome space of the X_n and Y_n, and if $\mathscr{L}X_0$ is absolutely continuous (with respect to Lebesgue measure), then $\mathscr{L}(X_n + Y_n) \to \mathscr{L}(X_0 + Y_0)$ in total variation.

1. Introduction. The set of finite signed measures on the measurable space $(\mathscr{X}, \mathscr{F})$ is a Banach space with the total variation norm $\|\cdot\|$, also called the L_1-norm, which is defined by $\|\tau\| = |\tau|(\mathscr{X})$, where $|\tau|$ denotes the variation measure of τ. In the sequel we shall be interested in the subspace of probability measures. For probability measures P and Q,

$$\|P - Q\| = 2\sup\{|P(B) - Q(B)|;\, B \in \mathscr{F}\} = \int_{\mathscr{X}} |f - g|\, d\lambda, \tag{1.1}$$

where f and g are density functions of P and Q, resp., with respect to a σ-finite measure λ dominating P and Q (e.g. $\frac{1}{2}(P + Q)$).

Let $X_0, X_1, X_2, \ldots$ be a sequence of random variables with values in $(\mathscr{X}, \mathscr{F})$. Define $P_n = \mathscr{L}X_n$ for $n = 0, 1, 2, \ldots$. We write $X_n \xrightarrow{\text{tv}} X_0$ if $\|P_n - P_0\| \to 0$. In the following $\mathscr{X}$ will be a metric space and $\mathscr{F}$ the σ-algebra generated by the open sets. We write $X_n \xrightarrow{\mathscr{D}} X_0$ if $P_n \xrightarrow{\text{w}} P_0$.

The total variation norm can be used for a number of reasons. We shall present some examples. Note that convergence in total variation is stronger than weak convergence, see (1.1). Sometimes Scheffé's theorem (see Theorem 3.3) is used in order to establish weak convergence, but in fact total variation convergence is shown. The total variation norm plays an important part in the derivations of so-called zero-one laws or equivalence—orthogonality dichotomies of products of probability measures. We refer to e.g. Kakutani (**1948**), Blum and Pathak (**1972**), Nemetz (**1975**), Sendler (**1975**), Hillion (**1976**) and Steerneman (**1983**).

Suppose $(\mathscr{X}, \mathscr{F}) = (\mathbf{R}^k, \mathscr{B}_k)$, where $\mathscr{B}_k$ denotes the σ-algebra of Borel-measurable subsets. The following results on weak convergence are well known. Let X_n and Y_n be independently distributed k-dimensional random vectors, $n = 0, 1, 2, \ldots$. If $X_n \xrightarrow{\mathscr{D}} X_0$ and $Y_n \xrightarrow{\mathscr{D}} Y_0$, then $X_n + Y_n \xrightarrow{\mathscr{D}} X_0 + Y_0$. If $c_n \in \mathbf{R}$, $n = 0, 1, 2, \ldots$, with $c_n \to c_0$, then

Received by the editors October 18, 1984 and, in revised form, February 21, 1985.
1980 *Mathematics Subject Classification*. Primary 60B10.
Key words and phrases. Total variation norm, weak convergence.
[1]This research was supported by a grant from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) awarded to Ton Steerneman.

$c_n X_n \xrightarrow{\mathscr{D}} c_0 x_0$. If $X_n \xrightarrow{\text{tv}} X_0$ and $Y_n \xrightarrow{\text{tv}} Y_0$, then it follows from the Propositions 3.1 and 3.2 that $X_n + Y_n \xrightarrow{\text{tv}} X_0 + Y_0$. The question arises: What happens if $X_n \xrightarrow{\text{tv}} X_0$, and $Y_n \xrightarrow{\mathscr{D}} Y_0$ instead of $Y_n \xrightarrow{\text{tv}} Y_0$? Can we conclude that $X_n + Y_n \xrightarrow{\text{tv}} X_0 + Y_0$ and $c_n X_n \xrightarrow{\text{tv}} c_0 X_0$?

In §2 we present a theorem which implies the desired results. §3 gives some preliminary and useful results. The proof of the theorem in §2 is given in §4.

2. The main results. Let X_n and Y_n be independently distributed random variables with outcomes in $(\mathbf{R}^k, \mathscr{B}_k)$ for $n = 0, 1, 2, \ldots$. Assume that $X_n \xrightarrow{\text{tv}} X_0$ and $Y_n \xrightarrow{\text{w}} Y_0$. In the Introduction we asked ourselves whether $X_n + Y_n \xrightarrow{\text{tv}} X_0 + Y_0$. This question can affirmatively be answered in case that $\mathscr{L}X_0$ is absolutely continuous (with respect to Lebesgue measure). Under this additional condition we also have $c_n X_n \xrightarrow{\text{tv}} c_0 X_0$ if $c_n \in \mathbf{R}$ for $n = 0, 1, 2, \ldots$ and $c_n \to c_0 \neq 0$.

In the first case we are interested in $T(X_n, Y_n) = X_n + Y_n$. In the second situation we are concerned with $T(X_n, c_n) = c_n X_n$. So, we become interested in the following more general problem. Let X_n and Y_n for $n = 0, 1, 2, \ldots$ be independent random variables with values in the respective metric spaces $\mathscr{X}$ and $\mathscr{Y}$. Let $T\colon \mathscr{X} \times \mathscr{Y} \to \mathscr{X}$ be a measurable mapping. Suppose that $X_n \xrightarrow{\text{tv}} X_0$ and $Y_n \xrightarrow{\text{w}} Y_0$. Can we give sufficient conditions for $T(X_n, Y_n) \xrightarrow{\text{tv}} T(X_0, Y_0)$? The next theorem presents an answer to this question. The two results mentioned in the first paragraph of this section follow from this theorem, but they can also be obtained as immediate consequences of two corollaries given at the end of this section.

THEOREM 2.1. *Let $\mathscr{X}$ be a complete and separable metric space and let $\mathscr{Y}$ be a metric space with respective Borel σ-algebras $\mathscr{F}$, $\mathscr{G}$. Suppose μ is a σ-finite measure on $(\mathscr{X}, \mathscr{F})$ and $\{T_y\colon \mathscr{X} \to \mathscr{X} \mid y \in \mathscr{Y}\}$ is a family of bijections. Let the following conditions be satisfied*:

(i) *the mapping $T\colon \mathscr{X} \times \mathscr{Y} \to \mathscr{X}$ defined by $T(x, y) = T_y(x)$ is measurable*;

(ii) *the mapping $\tilde{T}\colon \mathscr{X} \times \mathscr{Y} \to \mathscr{X}$ defined by $\tilde{T}(x, y) = T_y^{-1}(x)$ is measurable, and $\tilde{T}(x, y)$ is continuous in y for each $x \in \mathscr{X}$*;

(iii) *μT_y^{-1} is dominated by μ for any $y \in \mathscr{Y}$*;

(iv) *there exists a measurable function $h\colon \mathscr{X} \times \mathscr{Y} \to [0, \infty)$ such that*

(1) *$h(x, y) = (d\mu T_y^{-1}/d\mu)(x)$ a.e. $x[\mu]$, for any $y \in \mathscr{Y}$,*

(2) *$h(x, y)$ is continuous in y for any $x \in \mathscr{X}$, and*

(3) *$\sup\{h(x, y) \mid y \in \mathscr{Y},\ T_y^{-1}(x) \in K\} < \infty$ for each fixed $x \in \mathscr{X}$, $K \subset \mathscr{X}$, diameter $K < \infty$.*

If $\{P_n\}$, $\{Q_n\}$, $n = 0, 1, 2, \ldots$, are sequences of probability measures on $(\mathscr{X}, \mathscr{F})$, $(\mathscr{Y}, \mathscr{G})$, respectively, $P_0 \ll \mu$, $\|P_n - P_0\| \to 0$, and $Q_n \xrightarrow{\text{w}} Q_0$ as $n \to \infty$, then

$$\left\|(P_n \times Q_n)T^{-1} - (P_0 \times Q_0)T^{-1}\right\| \to 0 \quad \text{as } n \to \infty.$$

For the proof of this theorem we refer to §4. The theorem has two simple consequences.

COROLLARY 2.2. *Let $\mathscr{X}$ be a locally compact group satisfying the second axiom of countability with Borel σ-algebra $\mathscr{F}$, and let $\{\tau_n\}$, $n = 0,1,2,\ldots$, be a sequence of bimeasurable automorphisms of $\mathscr{X}$ such that $\tau_n \to \tau_0$ as $n \to \infty$, uniformly on compacta. If $\{P_n\}$, $n = 0,1,2,\ldots$, is a sequence of probability measures on $(\mathscr{X}, \mathscr{F})$, $\|P_n - P_0\| \to 0$ as $n \to \infty$, and P_0 is dominated by a left invariant Haar measure on $(\mathscr{X}, \mathscr{F})$, then $\|P_n\tau_n^{-1} - P_0\tau_0^{-1}\| \to 0$.*

PROOF. Note that $\mathscr{X}$ is a complete, separable and metrizable topological group. Let $\mathscr{Y}$ be the compact space defined by the image of the sequence $\{\tau_n\}$, $n = 0,1,2,\ldots$, and Q_n the Dirac measure at τ_n. □

COROLLARY 2.3. *Let $\mathscr{X}$ be a locally compact group satisfying the second axiom of countability with Borel σ-algebra $\mathscr{F}$, and let $\{P_n\}$, $\{Q_n\}$, $n = 0,1,2,\ldots$, be sequences of probability measures on $(\mathscr{X}, \mathscr{F})$ such that $\|P_n - P_0\| \to 0$ and $Q_n \xrightarrow{w} Q_0$ as $n \to \infty$. If P_0 is dominated by a right invariant (resp. left invariant) Haar measure on $(\mathscr{X}, \mathscr{F})$, then $\|P_n * Q_n - P_0 * Q_0\| \to 0$ (resp. $\|Q_n * P_n - Q_0 * P_0\| \to 0$) as $n \to \infty$.*

PROOF. Let $\mathscr{Y} = \mathscr{X}$ and define $T_y(x) = xy$. □

3. Preliminaries. From (1.1) we have the following result:

PROPOSITION 3.1. *Let $(\mathscr{X}, \mathscr{F})$ and $(\mathscr{Y}, \mathscr{G})$ be measurable spaces and $T\colon \mathscr{X} \to \mathscr{Y}$ be a measurable mapping. Let P and Q be probability measures on $(\mathscr{X}, \mathscr{F})$ and let PT^{-1} and QT^{-1} denote the induced probability measures on $(\mathscr{Y}, \mathscr{G})$, then $\|PT^{-1} - QT^{-1}\| \leqslant \|P - Q\|$.*

Sendler (**1975**) established the following result:

PROPOSITION 3.2. *For $i = 1,\ldots,n$ let $(\mathscr{X}_i, \mathscr{F}_i)$ be measurable spaces and let P_i and Q_i be probability measures on $(\mathscr{X}_i, \mathscr{F}_i)$, then $\|X_{i=1}^n P_i - X_{i=1}^n Q_i\| \leqslant \Sigma_{i=1}^n \|P_i - Q_i\|$.*

A very useful tool in establishing total variation convergence is Scheffé's theorem (see e.g. Billingsley (**1968**), pp. 223–224).

THEOREM 3.3. *Let $(\mathscr{X}, \mathscr{F})$ be a measurable space. Let $P_0, P_1, P_2, \ldots$ be a sequence of probability measures on $(\mathscr{X}, \mathscr{F})$ being dominated by the σ-finite measure μ. For $n = 0,1,2,\ldots$ let $p_n \in dP_n/d\mu$ be a density function of P_n with respect to μ. If $p_n \to p_0$ a.e. $[\mu]$, then $\|P_n - P_0\| \to 0$.*

4. Proof of Theorem 2.1. In $\mathscr{L}_1(\mathscr{X}, \mathscr{F}, \mu)$, bounded, continuous functions with support of finite diameter are dense. If f_0 is a version of the Radon-Nikodym derivative of P_0 with respect to μ, then choose for any $\varepsilon > 0$ a nonnegative, bounded, continuous function f_ε with support of finite diameter such that $\int f_\varepsilon \, d\mu = 1$ and

$$\int |f_\varepsilon - f_0| d\mu < \varepsilon. \tag{4.1}$$

Define

$$P^\varepsilon(E) = \int_E f_\varepsilon \, d\mu, \qquad E \in \mathscr{F}.$$

Now we have according to Propositions 3.1 and 3.2

$$\begin{aligned}
&\|(P_n \times Q_n)T^{-1} - (P_0 \times Q_0)T^{-1}\| \\
&\quad \leqslant \|(P_n \times Q_n)T^{-1} - (P_0 \times Q_n)T^{-1}\| + \|(P_0 \times Q_n)T^{-1} - (P^\varepsilon \times Q_n)T^{-1}\| \\
&\qquad + \|(P^\varepsilon \times Q_n)T^{-1} - (P^\varepsilon \times Q_0)T^{-1}\| + \|(P^\varepsilon \times Q_0)T^{-1} - (P_0 \times Q_0)T^{-1}\| \\
&\quad \leqslant \|P_n - P_0\| + \|P_0 - P^\varepsilon\| \\
&\qquad + \|(P^\varepsilon \times Q_n)T^{-1} - (P^\varepsilon \times Q_0)T^{-1}\| + \|P^\varepsilon - P_0\|.
\end{aligned}$$

By (1.1) and (4.1) we obtain

$$\begin{aligned}
(4.2)\qquad &\limsup_{n\to\infty} \|(P_n \times Q_n)T^{-1} - (P_0 \times Q_0)T^{-1}\| \\
&\qquad \leqslant 2\varepsilon + \limsup_{n\to\infty} \|(P^\varepsilon \times Q_n)T^{-1} - (P^\varepsilon \times Q_0)T^{-1}\|.
\end{aligned}$$

For any $E \in \mathscr{F}$, $n = 0, 1, 2, \ldots,$ and $\varepsilon > 0$, we derive

$$\begin{aligned}
(P^\varepsilon \times Q_n)T^{-1}(E) &= \int\int I_E\big(T_y(x)\big)\, dP^\varepsilon(x)\, dQ_n(y) \\
&= \int\int I_E\big(T_y(x)\big) f_\varepsilon(x)\, d\mu(x)\, dQ_n(y) \\
&= \int\int I_E(x) f_\varepsilon\big(T_y^{-1}(x)\big) h(x, y)\, dQ_n(y)\, d\mu(x) \\
&= \int_E g_n^\varepsilon(x)\, d\mu(x),
\end{aligned}$$

where I_E is the indicator function of E and

$$(4.3)\qquad g_n^\varepsilon(x) = \int f_\varepsilon\big(T_y^{-1}(x)\big) h(x, y)\, dQ_n(y).$$

So g_n^ε is a probability density function (with respect to μ) of $(P^\varepsilon \times Q_n)T^{-1}$. On account of the conditions (ii) and (iv) we have that $f_\varepsilon(T_y^{-1}(x))h(x, y)$ is bounded and continuous in y for any $x \in \mathscr{X}$. Since $Q_n \xrightarrow{\mathrm{w}} Q_0$ it now follows from (4.3) that

$$\lim_{n\to\infty} g_n^\varepsilon(x) = g_0^\varepsilon(x) \quad \text{for any } x \in \mathscr{X}.$$

By Scheffé's theorem

$$\lim_{n\to\infty} \|(P^\varepsilon \times Q_n)T^{-1} - (P^\varepsilon \times Q_0)T^{-1}\| = 0.$$

The proof is completed by using (4.2). □

ACKNOWLEDGMENTS. The authors would like to thank Dr. T. A. B. Snijders for his helpful suggestions, and a referee for his suggestions regarding the presentation.

REFERENCES

P. Billingsley (**1968**), *Convergence of probability measures*, Wiley, New York.

J. R. Blum and P. K. Pathak (**1972**), *A note on the zero-one law*, Ann. Math. Statist. **43**, 1008–1009.

A. Hillion (**1976**), *Sur l'intégrale Hellinger et la separation asymptotique*, C. R. Acad. Sci. Paris Sér. A **283**, 61–64.

S. Kakutani (**1948**), *On equivalence of infinite product measures*, Ann. of Math. (2) **49**, 214–224.

T. Nemetz (**1975**), *Equivalence-orthogonality dichotomies of probability measures*, Colloquium Mathematica Societatis János Bolyai, 11, Limit Theorems in Probability Theory, Keszthely, 1974, Hungary; North-Holland, Amsterdam.

W. Sendler (**1975**), *A note on the proof of the zero-one law of Blum and Pathak*, Ann. Probab. **3**, 1055–1058.

A. G. M. Steerneman (**1983**), *On the total variation and Hellinger distance between signed measures; an application to product measures*, Proc. Amer. Math. Soc. **88**, 684–688.

Indian Statistical Institute, Delhi Centre, 7, S. J. S. Sansanwal Marg, New Delhi 110016, India

Econometrics Institute, University of Groningen, P. O. Box 800, 9700 AV Groningen, The Netherlands