PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 88, Number 4, August 1983

ON THE TOTAL VARIATION AND HELLINGER DISTANCE BETWEEN SIGNED MEASURES; AN APPLICATION TO PRODUCT MEASURES[1]

TON STEERNEMAN

ABSTRACT. Firstly, the Hellinger metric on the set of probability measures on a measurable space is extended to the set of signed measures. An inequality between total variation and Hellinger metric due to Kraft is generalized to the case of signed measures. The inequality is used in order to derive a lower estimate concerning the total variation distance between products of signed measures. The lower bound depends on the total variation norms of the signed measures and the total variation distances between the total variation measures of the single components.

1. Introduction. The set of finite signed measures on the measurable space $(\mathfrak{X}, \mathfrak{F})$ will be denoted by $\mathfrak{M} = \mathfrak{M}(\mathfrak{X}, \mathfrak{F})$ and $\mathfrak{M}_1 = \mathfrak{M}_1(\mathfrak{X}, \mathfrak{F})$ will be the set of probability measures. $\mathfrak{M}$ is a Banach lattice with respect to the so-called total variation norm $\|\cdot\|$. $\mathfrak{M}$ is distinguished by the extra property that $\|\mu + \nu\| = \|\mu\| + \|\nu\|$ for $\mu, \nu \geqslant 0$ (see Yosida [**7**, pp. 369–370]). Apart from the total variation metric d_v another metric is defined in §2 by extending the Hellinger d_H on $\mathfrak{M}_1$. Kraft [**4**, Lemma 1] provides the following important inequality:

$$d_H^2(P, Q) \leqslant d_v(P, Q) \leqslant 2 d_H(P, Q) \quad \forall P, Q \in \mathfrak{M}_1. \tag{1.1}$$

In §2 this inequality is extended to $\mathfrak{M}$. The result obtained will be applied in §3, where the total variation distance between two products of finite signed measures is estimated. For $i = 1, \ldots, k$ let $(\mathfrak{X}_i, \mathfrak{F}_i)$ be measurable spaces, $\mu_i \in \mathfrak{M}(\mathfrak{X}_i, \mathfrak{F}_i)$, then $\underline{X}_{i=1}^k \mu_i$ denotes the product measure on $(\underline{X}_{i=1}^k \mathfrak{X}_i, \underline{X}_{i=1}^k \mathfrak{F}_i)$. When $(\mathfrak{X}_i, \mathfrak{F}_i) = (\mathfrak{X}, \mathfrak{F})$ and $\mu_i = \mu$ we write μ^k and $(\mathfrak{X}^k, \mathfrak{F}^k)$. Hoeffding and Wolfowitz [**3**, (4.4) and (4.5)] imply that

$$d_v(P, Q) \leqslant d_v(P^k, Q^k) \leqslant k d_v(P, Q) \quad \forall P, Q \in \mathfrak{M}_1. \tag{1.2}$$

The upper estimate is extended by Blum and Pathak [**2**, Lemma 1.3], and Sendler [**6**, Lemma 2.1] to the product of k not necessarily identical finite signed measures. An upper bound for a situation with nonidentical finite signed measures is given by Reiss [**5**, Theorem 2.1], who uses the sup-metric instead of the total variation metric. A lower estimate for nonidentical probability measures was found by Behnen and Neuhaus [**1**] (see proof of proposition, pp. 1351–1352). In §3 we shall derive upper and lower bounds for $d_v(\underline{X}_{i=1}^k \mu_i, \underline{X}_{i=1}^k \nu_i)$, where $\mu_i, \nu_i \in \mathfrak{M}(\mathfrak{X}_i, \mathfrak{F}_i)$ for $i = 1, \ldots, k$.

Received by the editors May 21, 1982.
1980 *Mathematics Subject Classification*. Primary 46E27; Secondary 60E15.
[1]This research was supported by a grant from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

2. Total variation and Hellinger metric. Let $\|\cdot\|$ denote the total variation norm on $\mathfrak{M}$, that is $\|\mu\| = |\mu|(\mathfrak{X})$ where $|\mu|$ denotes the total variation measure of μ. For $\mu, \nu \in \mathfrak{M}$ let λ be a σ-finite measure on $(\mathfrak{X}, \mathfrak{F})$ such that $\mu, \nu \ll \lambda$ and let $f \in du/d\lambda$, $g \in d\nu/d\lambda$, then

$$d_v(\mu, \nu) = \int |f - g|\, d\lambda. \tag{2.1}$$

Note that $|\|\mu\| - \|\nu\|| \leqslant d_v(\mu, \nu) \leqslant \|\mu\| + \|\nu\|$. We may assume that λ is a finite measure, e.g. $\lambda = |\mu| + |\nu|$. It can easily be seen that

$$d_v(\mu, \nu) = 2\sup\{|\mu(A) - \nu(A)|: A \in \mathfrak{F}\} \quad \text{if } \mu(\mathfrak{X}) = \nu(\mathfrak{X}). \tag{2.2}$$

We define the (generalized) Hellinger affinity between μ and ν by

$$\rho(\mu, \nu) = \int_A |fg|^{1/2}\, d\lambda, \qquad A = \{fg > 0\}. \tag{2.3}$$

Note that $\rho(\mu, \nu)$ does not depend on the particular choice of λ. By Hölder's inequality:

$$0 \leqslant \rho(\mu, \nu) \leqslant \|\mu\|^{1/2}\|\nu\|^{1/2}.$$

The Hellinger distance on $\mathfrak{M}$ is defined by

$$d_H(\mu, \nu) = [\|\mu\| + \|\nu\| - 2\rho(\mu, \nu)]^{1/2}. \tag{2.4}$$

Note that $|\|\mu\|^{1/2} - \|\nu\|^{1/2}| \leqslant d_H(\mu, \nu) \leqslant (\|\mu\| + \|\nu\|)^{1/2}$.

In order to show that d_H is a metric the following alternative expression turns out to be useful

$$d_H^2(\mu, \nu) = \int \left(f_+^{1/2} - g_+^{1/2}\right)^2 d\lambda + \int \left(f_-^{1/2} - g_-^{1/2}\right)^2 d\lambda, \tag{2.5}$$

where $x_+ = \max(x, 0)$ and $x_- = -\min(x, 0)$. The triangle inequality can be verified by using Minkowski's inequality and

$$(ab)^{1/2} + (cd)^{1/2} \leqslant (a + c)^{1/2}(b + d)^{1/2}, \qquad a, b, c, d \geqslant 0. \tag{2.6}$$

The following result is a generalisation of Kraft's inequality given in (1.1).

THEOREM 2.1. *For* $\mu, \nu \in \mathfrak{M}(\mathfrak{X}, \mathfrak{F})$

$$\begin{aligned} d_H^2(\mu, \nu) \leqslant d_v(\mu, \nu) &\leqslant \left[(\|\mu\| + \|\nu\|)^2 - 4\rho^2(\mu, \nu)\right]^{1/2} \\ &\leqslant \left(\|\mu\|^{1/2} + \|\nu\|^{1/2}\right) d_H(\mu, \nu). \end{aligned} \tag{2.7}$$

PROOF. Let λ be a σ-finite measure on $(\mathfrak{X}, \mathfrak{F})$ with $\mu, \nu \ll \lambda$, and $f \in d\mu/d\lambda$, $g \in d\nu/d\lambda$. According to (2.5) and the inequality

$$(a^{1/2} - b^{1/2})^2 \leqslant |a - b|, \qquad a, b \geqslant 0, \tag{2.8}$$

it can be written

$$\begin{aligned} d_H^2(\mu, \nu) &\leqslant \int |f_+ - g_+|\, d\lambda + \int |f_- - g_-|\, d\lambda \\ &= \int |f - g|\, d\lambda = d_v(\mu, \nu), \end{aligned}$$

because $|x_+ - y_+| + |x_- - y_-| = |x - y|$ for any $x, y \in \mathbf{R}$.

The second part of inequality (2.7) can be established as follows. Defining $A = \{fg > 0\}$, $B = \{fg \leqslant 0\}$, Hölder's inequality provides

$$d_v(\mu, \nu) = \int |f - g|\, d\lambda = \int_A \big||f| - |g|\big|\, d\lambda + \int_B (|f| + |g|)\, d\lambda$$

$$= \int_A \big||f|^{1/2} - |g|^{1/2}\big| \cdot \left(|f|^{1/2} + |g|^{1/2}\right) d\lambda + \int_B (|f| + |g|)\, d\lambda$$

$$\leqslant \left[\int_A \left(|f|^{1/2} - |g|^{1/2}\right)^2 d\lambda\right]^{1/2} \left[\int_A \left(|f|^{1/2} + |g|^{1/2}\right)^2 d\lambda\right]^{1/2} + \int_B (|f| + |g|)\, d\lambda.$$

Next by applying inequality (2.6) it follows

$$d_v^2(\mu, \nu) \leqslant \left[\int_A \left(|f|^{1/2} - |g|^{1/2}\right)^2 d\lambda + \int_B (|f| + |g|)\, d\lambda\right]$$

$$\times \left[\int_A \left(|f|^{1/2} + |g|^{1/2}\right) d\lambda + \int_B (|f| + |g|)\, d\lambda\right]$$

$$= \left[\|\mu\| + \|\nu\| - 2\rho(\mu, \nu)\right]\left[\|\mu\| + \|\nu\| + 2\rho(\mu, \nu)\right]$$

$$= \left(\|\mu\| + \|\nu\|\right)^2 - 4\rho^2(\mu, \nu).$$

The last inequality in (2.7) is found if, returning to the last but one formula, the first factor is written as $d_H^2(\mu, \nu)$ and the second is estimated by $(\|\mu\|^{1/2} + \|\nu\|^{1/2})^2$. □

Remark 2.1. Formula (1.1) keeps valid for signed measures with total variation norms equal to one, due to the definitions (2.3) and (2.4).

Remark 2.2. The Hellinger metric does not fit well into the linear structure of $\mathfrak{M}$. However, it is often easier to calculate in practice. The total variation and Hellinger metric induce the same topology on $\mathfrak{M}$. Moreover, they induce the same uniformity structure on $\mathfrak{M}_1$; this is not true for $\mathfrak{M}$ in general.

3. Approximating the total variation distance between products of signed measures.

Theorem 3.1. *Let* $\mu_i, \nu_i \in \mathfrak{M}(\mathfrak{X}_i, \mathfrak{F}_i)$, $i = 1, \dots, k$. *Then*

$$\text{(3.1)} \quad \prod_{i=1}^{k} \|\mu_i\| + \prod_{i=1}^{k} \|\nu_i\| - 2^{1-k} \prod_{i=1}^{k} \left[\left(\|\mu_i\| + \|\nu_i\|\right)^2 - d_v^2(|\mu_i|, |\nu_i|)\right]^{1/2}$$

$$\leqslant d_v\left(\underline{X}_{i=1}^{k} \mu_i, \underline{X}_{i=1}^{k} \nu_i\right) \leqslant \sum_{i=1}^{k} \left(\prod_{j=1}^{i-1} \|\mu_j\|\right)\left(\prod_{j=i+1}^{k} \|\nu_j\|\right) d_v(\mu_i, \nu_i).$$

Products with empty index sets are defined to be equal to one.

Proof. For $i = 1, \dots, k$ let λ_i be a σ-finite measure on $(\mathfrak{X}_i, \mathfrak{F}_i)$ with $\mu_i, \nu_i \ll \lambda_i$ and $f_i \in d\mu_i/d\lambda_i$, $g_i \in d\nu_i/d\lambda_i$. Using the formula

$$\prod_{j=1}^{n} x_j - \prod_{j=1}^{n} y_j = \sum_{i=1}^{n} \left(\prod_{j=1}^{i-1} x_j\right)\left(\prod_{j=i+1}^{n} y_j\right)(x_i - y_i)$$

and the triangle inequality the second part of (3.1) easily follows. Remark that

$$\rho\left(\underline{X}_{i=1}^{k} \mu_i, \underline{X}_{i=1}^{k} \nu_i\right) \leqslant \prod_{i=1}^{k} \rho(|\mu_i|, |\nu_i|).$$

By applying Theorem 2.1 two times we get

$$\prod_{i=1}^{k}\|\mu_i\| + \prod_{i=1}^{k}\|\nu_i\| - d_v\left(\underline{X}_{i=1}^{k}\mu_i, \underline{X}_{i=1}^{k}\nu_i\right) \leqslant 2\prod_{i=1}^{k}\rho(|\mu_i|,|\nu_i|)$$

$$\leqslant 2^{1-k}\prod_{i=1}^{k}\left[(\|\mu_i\| + \|\nu_i\|)^2 - d_v^2(|\mu_i|,|\nu_i|)\right]^{1/2}. \quad \square$$

Using the inequality $1 - x \leqslant \exp(-x)$ we get

COROLLARY 3.1. *Let* $\mu_i, \nu_i \in \mathfrak{M}(\mathfrak{X}_i, \mathfrak{F}_i)$, $i = 1,\dots, k$. *Then*

(3.2)

$$d_v\left(\underline{X}_{i=1}^{k}\mu_i, \underline{X}_{i=1}^{k}\nu_i\right) \geqslant \prod_{i=1}^{k}\|\mu_i\| + \prod_{i=1}^{k}\|\nu_i\| + \\ - 2^{1-k}\left[\prod_{i=1}^{k}(\|\mu_i\| + \|\nu_i\|)\right]\exp\left[-\frac{1}{2}\sum_{i=1}^{k}(\|\mu_i\| + \|\nu_i\|)^{-2}d_v^2(|\mu_i|,|\nu_i|)\right].$$

An immediate consequence of Corollary 3.1 is

COROLLARY 3.2. *Let* $P_i, Q_i \in \mathfrak{M}_1(\mathfrak{X}_i, \mathfrak{F}_i)$, $i = 1,\dots, k$, *then*

$$2 - 2\exp\left(-8^{-1}\sum_{i=1}^{k}d_v^2(P_i, Q_i)\right) \leqslant d_v\left(\underline{X}_{i=1}^{k}P_i, \underline{X}_{i=1}^{k}Q_i\right) \leqslant \sum_{i=1}^{k}d_v(P_i, Q_i). \tag{3.3}$$

This result is the same as (1.2) in Reiss [5] because of (2.2).

The following theorem provides another lower bound for $d_v(\underline{X}_{i=1}^{k}\mu_i, \underline{X}_{i=1}^{k}\nu_i)$.

THEOREM 3.2. *Let* $\mu_i, \nu_i \in \mathfrak{M}(\mathfrak{X}_i, \mathfrak{F}_i)$, $i = 1,\dots, k$. *Then*

$$d_v\left(\underline{X}_{i=1}^{k}\mu_i, \underline{X}_{i=1}^{k}\nu_i\right) \geqslant \prod_{i=1}^{k}\|\mu_i\| + \prod_{i=1}^{k}\|\nu_i\| + \\ - 2\left[\prod_{i=1}^{k}\|\mu_i\|\cdot\|\nu_i\|\right]^{1/2}\exp\left(-8^{-1}\sum_{i=1}^{k}d_v^2(\tilde{\mu}_i, \tilde{\nu}_i)\right),$$

where $\tilde{\mu}_i = \|\mu_i\|^{-1}|\mu_i|$, $\tilde{\nu}_i = \|\nu_i\|^{-1}|\nu_i|$.

PROOF. According to Theorem 2.1 we have

$$\prod_{i=1}^{k}\|\mu_i\| + \prod_{i=1}^{k}\|\nu_i\| - d_v\left(\underline{X}_{i=1}^{k}\mu_i, \underline{X}_{i=1}^{k}\nu_i\right) \leqslant 2\prod_{i=1}^{k}\|\mu_i\|^{1/2}\|\nu_i\|^{1/2}\rho(\tilde{\mu}_i, \tilde{\nu}_i)$$

$$\leqslant 2^{1-k}\prod_{i=1}^{k}\|\mu_i\|^{1/2}\|\nu_i\|^{1/2}\left[4 - d_v(\tilde{\mu}_i, \tilde{\nu}_i)\right]^{1/2}$$

$$\leqslant 2\left[\prod_{i=1}^{k}\|\mu_i\|\|\nu_i\|\right]^{1/2}\exp\left(-8^{-1}\sum_{i=1}^{k}d_v^2(\tilde{\mu}_i, \tilde{\nu}_i)\right). \quad \square$$

ACKNOWLEDGMENT. I wish to thank Professor Dr. W. Schaafsma and Professor Dr. A. J. Stam for their careful reading of earlier versions of this note.

REFERENCES

1. K. Behnen and G. Neuhaus, *A central limit theorem under contiguous alternatives*, Ann. Statist. **3** (1975), 1349–1353.

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

3. W. Hoeffding and J. Wolfowitz, *Distinguishability of sets of distributions*, Ann. Math. Statist. **29** (1958), 700–718.

4. C. Kraft, *Some conditions for consistency and uniform consistency of statistical procedures*, Univ. California Publications in Statist., Vol. 2, 1955, pp. 125–142.

5. R.-D. Reiss, *Approximation of product measures with an application to order statistics*, Ann. Probability **9** (1981), 335–341.

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

7. K. Yosida, *Functional analysis*, Springer-Verlag, Berlin and New York, 1978.

ECONOMETRICS INSTITUTE, STATE UNIVERSITY GRONINGEN, P. O. BOX 800, 9700 AV GRONINGEN, THE NETHERLANDS