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

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 \((\mathcal{X}, \mathcal{F})\) will be denoted by \(\mathcal{M} = \mathcal{M}(\mathcal{X}, \mathcal{F})\) and \(\mathcal{M}_1 = \mathcal{M}_1(\mathcal{X}, \mathcal{F})\) will be the set of probability measures. \(\mathcal{M}\) is a Banach lattice with respect to the so-called total variation norm \(\| \cdot \|\). \(\mathcal{M}_1\) is distinguished by the extra property that \(\| \mu + \nu \| = \| \mu \| + \| \nu \|\) for \(\mu, \nu \geq 0\) (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 \(\mathcal{M}_1\). Kraft [4, Lemma 1] provides the following important inequality:

\[
(1.1) \quad d_H^2(\mu, \nu) \leq d_v(\mu, \nu) \leq 2d_H(\mu, \nu) \quad \forall \mu, \nu \in \mathcal{M}_1.
\]

In §2 this inequality is extended to \(\mathcal{M}\). 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 \((\mathcal{X}_i, \mathcal{F}_i)\) be measurable spaces, \(\mu_i \in \mathcal{M}(\mathcal{X}_i, \mathcal{F}_i)\), then \(X_i = \prod_{i=1}^k \mathcal{X}_i\) denotes the product measure on \((X_i, \mathcal{F}_i, \mathcal{F}_i)\). When \((\mathcal{X}_i, \mathcal{F}_i) = (\mathcal{X}, \mathcal{F})\) and \(\mu_i = \mu\) we write \(\mu^k\) and \((\mathcal{X}^k, \mathcal{F}^k)\). Hoeffding and Wolfowitz [3, (4.4) and (4.5)] imply that

\[
(1.2) \quad d_v(\mu, \nu) \leq d_v(\mu^k, \nu^k) \leq kd_v(\mu, \nu) \quad \forall \mu, \nu \in \mathcal{M}_1.
\]

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 \(d_v(X_i^k \mu_i, X_i^k \nu_i)\), where \(\mu_i, \nu_i \in \mathcal{M}(\mathcal{X}_i, \mathcal{F}_i)\) for \(i = 1, \ldots, k\).
2. Total variation and Hellinger metric. Let $\| \cdot \|$ denote the total variation norm on $\mathcal{M}$, that is $\| \mu \| = |\mu| (\mathcal{F})$ where $|\mu|$ denotes the total variation measure of $\mu$. For $\mu, \nu \in \mathcal{M}$ let $\lambda$ be a $\sigma$-finite measure on $(\mathcal{F}, \mathcal{G})$ such that $\mu, \nu \ll \lambda$ and let $f \in du/d\lambda, g \in dv/d\lambda$, then

\begin{equation}
(2.1) \quad d_v (\mu, \nu) = \int |f - g| d\lambda.
\end{equation}

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

\begin{equation}
(2.2) \quad d_v (\mu, \nu) = 2 \sup \{ |\mu (A) - \nu (A)| ; A \in \mathcal{G} \} \quad \text{if} \quad \mu (\mathcal{F}) = \nu (\mathcal{F}).
\end{equation}

We define the (generalized) Hellinger affinity between $\mu$ and $\nu$ by

\begin{equation}
(2.3) \quad \rho (\mu, \nu) = \int_A |fg|^{1/2} d\lambda, \quad A = \{ fg > 0 \}.
\end{equation}

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

\begin{equation}
0 \leq \rho (\mu, \nu) \leq \| \mu \|^{1/2} \| \nu \|^{1/2}.
\end{equation}

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

\begin{equation}
(2.4) \quad d_H (\mu, \nu) = [\| \mu \| + \| \nu \| - 2 \rho (\mu, \nu)]^{1/2}.
\end{equation}

Note that $\| \| \mu \| - \| \nu \| \| \leq d_H (\mu, \nu) \leq (\| \mu \| + \| \nu \|) \| \leq 2$.

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

\begin{equation}
(2.5) \quad d_H^2 (\mu, \nu) = \int (f^{+2} - g^{+2})^2 d\lambda + \int (f^{-2} - g^{-2})^2 d\lambda,
\end{equation}

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

\begin{equation}
(2.6) \quad (ab)^{1/2} + (cd)^{1/2} \leq (a + c)^{1/2} (b + d)^{1/2}, \quad a, b, c, d \geq 0.
\end{equation}

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

**Theorem 2.1.** For $\mu, \nu \in \mathcal{M} (\mathcal{F}, \mathcal{G})$

\begin{equation}
(2.7) \quad d_H^2 (\mu, \nu) \leq d_v (\mu, \nu) \leq \left[ (\| \mu \| + \| \nu \|)^2 - 4 \rho^2 (\mu, \nu) \right]^{1/2}
\end{equation}

\begin{equation}
\leq \left( \| \mu \|^{1/2} + \| \nu \|^{1/2} \right) d_H (\mu, \nu).
\end{equation}

**Proof.** Let $\lambda$ be a $\sigma$-finite measure on $(\mathcal{F}, \mathcal{G})$ with $\mu, \nu \ll \lambda$, and $f \in du/d\lambda, g \in dv/d\lambda$. According to (2.5) and the inequality

\begin{equation}
(2.8) \quad (a^{1/2} - b^{1/2})^2 \leq |a - b|, \quad a, b \geq 0,
\end{equation}

it can be written

\begin{equation}
(2.9) \quad d_H^2 (\mu, \nu) = \int |f_+ - g_+| d\lambda + \int |f_- - g_-| d\lambda
\end{equation}

\begin{equation}
= \int |f - g| d\lambda = d_v (\mu, \nu),
\end{equation}

because $|x^+ - y^+| + |x^- - y^-| = |x - y|$ for any $x, y \in \mathbb{R}$.
The second part of inequality (2.7) can be established as follows. Defining

\[ A = \{fg > 0\}, \quad B = \{fg \leq 0\}, \]

Hölder’s inequality provides

\[
d_{\nu}(\mu, \nu) = \int |f - g| \, d\lambda = \int_A |f| \, d\lambda + \int_B (|f| + |g|) \, d\lambda
\]

\[
= \int_A |f|^{1/2} - |g|^{1/2} \cdot (|f|^{1/2} + |g|^{1/2}) \, d\lambda + \int_B (|f| + |g|) \, d\lambda
\]

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

Next by applying inequality (2.6) it follows

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

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

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

\[
= (||\mu|| + ||\nu||)^2 - 4\rho^2(\mu, \nu).
\]

The last inequality in (2.7) is found if, returning to the last but one formula, the first
factor is written as \(d_{\nu}^2(\mu, \nu)\) 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 \(\mathcal{M}\). However, it is often easier to calculate in practice. The total variation and Hellinger metric induce the same topology on \(\mathcal{M}\). Moreover, they induce the same uniformity structure on \(\mathcal{M}_i\); this is not true for \(\mathcal{M}\) in general.

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

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

\[
(3.1) \quad \prod_{i=1}^k ||\mu_i|| + \prod_{i=1}^k ||\nu_i|| - 2^{1-k} \prod_{i=1}^k \left[ (||\mu_i|| + ||\nu_i||)^2 - d_{\nu}^2(\mu_i, \nu_i) \right]^{1/2} \leq d_{\nu}(X^k_{i=1}, X^k_{i=1}) \leq \prod_{i=1}^k \left[ \prod_{j=1}^{i-1} ||\mu_j|| \right] \left( \prod_{j=i+1}^k ||\nu_j|| \right) d_{\nu}(\mu_i, \nu_i).
\]

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

**Proof.** For \(i = 1, \ldots, k\) let \(\lambda_i\) be a \(\sigma\)-finite measure on \((\mathcal{X}_i, \mathcal{F}_i)\) 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

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

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

\[
\rho(\frac{X^k_{i=1}}{X^k_{i=1}}, \frac{X^k_{i=1}}{X^k_{i=1}}) \leq \prod_{i=1}^k \rho(||\mu_i||, ||\nu_i||).
\]
By applying Theorem 2.1 two times we get
\[
\prod_{i=1}^{k} \|\mu_i\| + \prod_{i=1}^{k} \|\nu_i\| - d_v\left(\prod_{i=1}^{k} \mu_i, \prod_{i=1}^{k} \nu_i\right) \leq 2 \prod_{i=1}^{k} \rho(\mu_i, \nu_i)
\]
\[
\leq 2^{1-k} \prod_{i=1}^{k} \left((\|\mu_i\| + \|\nu_i\|)^2 - d_v^2(\mu_i, \nu_i)\right)^{1/2}.
\]

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

**Corollary 3.1.** Let \(\mu_i, \nu_i \in \mathcal{M}(\tilde{x}_i, \tilde{\nu}_i), i = 1, \ldots, k\). Then

(3.2)
\[
d_v\left(\prod_{i=1}^{k} \mu_i, \prod_{i=1}^{k} \nu_i\right) \geq \prod_{i=1}^{k} \|\mu_i\| + \prod_{i=1}^{k} \|\nu_i\| + \\
- 2^{1-k} \prod_{i=1}^{k} \left((\|\mu_i\| + \|\nu_i\|)\right) \exp\left(-\frac{1}{2} \sum_{i=1}^{k} \frac{1}{(\|\mu_i\| + \|\nu_i\|)^2} d_v^2(\mu_i, \nu_i)\right).
\]

An immediate consequence of Corollary 3.1 is

**Corollary 3.2.** Let \(P_i, Q_i \in \mathcal{M}(\tilde{x}_i, \tilde{\nu}_i), i = 1, \ldots, k\), then

(3.3)
\[
2 - 2 \exp\left(-8^{-1} \sum_{i=1}^{k} d_v^2(P_i, Q_i)\right) \leq d_v\left(\prod_{i=1}^{k} P_i, \prod_{i=1}^{k} Q_i\right) \leq \sum_{i=1}^{k} d_v(P_i, Q_i).
\]

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

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

**Theorem 3.2.** Let \(\mu_i, \nu_i \in \mathcal{M}(\tilde{x}_i, \tilde{\nu}_i), i = 1, \ldots, k\). Then

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

where \(\tilde{\nu}_i = \|\mu_i\|^{-1/2} \mu_i, \tilde{\nu}_i = \|\nu_i\|^{-1/2} \nu_i\).

**Proof.** According to Theorem 2.1 we have

\[
\prod_{i=1}^{k} \|\mu_i\| + \prod_{i=1}^{k} \|\nu_i\| - d_v\left(\prod_{i=1}^{k} \mu_i, \prod_{i=1}^{k} \nu_i\right) \leq 2 \prod_{i=1}^{k} \|\mu_i\|^{1/2} \|\nu_i\|^{1/2} \rho(\tilde{\mu}_i, \tilde{\nu}_i)
\]
\[
\leq 2^{1-k} \prod_{i=1}^{k} \left(\|\mu_i\|^{1/2} \|\nu_i\|^{1/2} \left[4 - d_v(\tilde{\mu}_i, \tilde{\nu}_i)\right]\right)^{1/2}
\]
\[
\leq 2 \left[\prod_{i=1}^{k} \|\mu_i\| \|\nu_i\|\right]^{1/2} \exp\left(-8^{-1} \sum_{i=1}^{k} d_v^2(\tilde{\mu}_i, \tilde{\nu}_i)\right). \quad \square
\]

**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